Sea Ice Concentration Estimation: Using Passive Microwave and SAR Data with Fully Convolutional Networks

by

Keerthijan Radhakrishnan

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Applied Science
in
Systems Design Engineering

Waterloo, Ontario, Canada, 2020

Author's Declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

Sea ice concentration is of great interest to ship navigators and scientists who require regional ice cover understanding. Passive microwave data and image analysis charts are typically used to estimate ice concentration, but these have limitations. Estimates obtained from passive microwave data have coarse spatial resolution, may be biased due to weather filters that reduce atmospheric contamination, and often perform poorly in marginal ice zones. Image analysis charts are not as precise and subjective to analyst interpretation.

Synthetic aperture radar (SAR) images are finer resolution satellite images that can be used to observe oceans. However, the complex interactions between the SAR signal and water and ice make it a difficult process to estimate sea ice concentration. Previous studies have found that deep learning is a viable avenue to estimate ice concentration from SAR images. In these studies, convolutional neural networks (CNNs) have been successful due to their ability to learn spatial features in their convolutional layers.

To overcome the shortcomings of ice concentration estimation, we have uniquely implemented a U-net with SAR images as inputs and use estimates obtained from passive microwave data as training labels. The U-net, due to not being sensitive to patch size, is shown to be an improvement over the CNN models used in previous studies. Data augmentation and a mean absolute error (L1) loss function were applied as well as a curriculum learning method that introduces more open water and consolidated ice regions before incorporating marginal ice regions. The key objectives of this study are (a) to overcome shortcomings of using passive microwave data for model training and (b) to improve ice concentration predictions in marginal ice zones.

Evaluating with image analysis charts, a mean absolute error of 7.18% is achieved, which is lower than errors associated with estimation algorithms using passive microwave data alone. Through qualitative analysis, we also show instances where our proposed model has more precise estimates in the marginal ice zones than traditionally used passive microwave data based ice concentration retrieval algorithms.

In this thesis, we also evaluate our proposed model to scenes from a higher latitude, which contain different ice types, to evaluate the ability of the model to extend to different scenes. It was found that model performance is worse in when extending to the new region. This suggests that the model was unable to learn features required to make accurate predictions on this dataset. Model performance may be improved if similar regions are included for training.

Lastly, we evaluate the effects of using training labels from different passive microwave data based retrieval algorithms. We show that model performance is affected by the training labels used and using different training labels have unique benefits and shortcomings.

To capitalize on benefits exclusive to specific training labels, we also attempt a method of staged training where the model is trained with one set of training labels before being trained with another set of training labels. We found that predictions made from models using staged training has qualities of the individual training labels used to train it. We also found that the final set of training labels used in staged training has the strongest impact on model predictions. From this experiment, we show that through staged training, we can teach deep learning models important information exclusive to different sets.

Acknowledgements

First and foremost, I would like to thank my supervisors, professor Andrea Scott and professor David Clausi, for their continuous help over the last two years. Thank you for your support and patience as I grow to be a researcher.

I would also like to thank my thesis committee members, Andrea Scott, David Clausi, and Kumaraswamy Ponnambalam from Systems Design Engineering and Grant Gunn from Geography and Environmental Management at the University of Waterloo. Thank you for taking time from your busy schedules to review my thesis.

More thanks to the members of the Vision and Image Processing Lab and Remote Sensing Research Group at the University of Waterloo. Thank you for giving me exposure to many interesting topics through weekly meetings and friendly discussions.

Lastly, I would like to thank my friends and family for providing support through this journey.

Thank you.

Dedication

This is dedicated to my family.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation

Sea ice concentration is calculated as a numeric value between zero and one, defined as the total area of ice in a specified region divided by the total area of that region. Navigators are interested in sea ice concentration because routes with icy conditions can be dangerous and time consuming. Vessels often get stuck in ice, which delays delivery time and expends fuel, with negative economic and environmental consequences. Captains of ice-class vessels have indicated that better knowledge of openings in the ice cover (leads) is important for safety [9]. For climate scientists, more accurate ice concentration information enables better models and ability to predict climate change occurrences [16].

Synthetic aperture radar (SAR) sensors and passive microwave sensors are more often used to monitor sea ice in arctic regions than optical sensors because they are less affected by cloud cover and do not depend on solar illumination. Data provided from these sensors can also be used to obtain ice concentration estimates. Images acquired from passive microwave sensors have a lower spatial resolution than images acquired from a SAR sensor and can be affected by atmosphere, but it depends on the band [2]. There are numerous algorithms using passive microwave data for ice concentration retrieval [23]. Algorithms using lower frequencies have a lower spatial resolution while higher frequency algorithms have higher spatial resolution, but they are more affected by atmospheric moisture. Developing a model to automate ice concentration estimates from SAR imagery is desirable because spatially precise estimates can be made.

For example, the smallest instantaneous field of view (IFOV) available from present passive microwave sensors is 3-5 km at 89 GHz, for which the atmospheric contamination

in the marginal ice zone is significant. Figure 1.1 shows a region where the shortcomings of sea ice concentration estimates obtained from passive microwave data is visible. The first estimation algorithm (retrieved from the ARTIST Sea Ice algorithm [61]) estimates consolidated ice regions in the marginal ice zone as lower ice concentration and the second estimation algorithm (retrieved from the NASA Team 2 algorithm [41]) lacks spatial details. The present study explores the possibility of improved ice concentration retrievals in the marginal ice zones through the use of SAR data. Although there are benefits to using SAR data to retrieve ice concentration estimates, it is a difficult task because of the complexity of the interaction of the SAR signal with water and ice. The backscatter of a SAR signal is dependent on the imaging geometry and surface conditions. Open water that is calm generally has a lower intensity than ice. Rough water conditions from wind and ocean currents can increase the backscatter. Previous studies have shown that deep learning models are a suitable choice for estimating ice concentration from SAR images because they can learn backscatter patterns of ice and water [64, 14].

Image analysis charts, which are manual analyses of SAR images carried out by trained ice experts, provide ice concentration estimates, but also have some shortcomings. First, with the increasing volumes of SAR images available, automated methods for ice concentration estimation from SAR images would be ideal. Moreover, the precision of the image analysis charts is no more than 10% because ice concentration estimates are given in increments of 0.1. Furthermore, image analysis charts provide a single label for a large spatial region with homogeneous ice characteristics. The ice concentration at a specific point may be different than the label given to the region. Furthermore, the image analysis charts have also been known to have bias due subjectivity of the ice experts [42].

1.2 Thesis Objectives

The goal of this project is to utilize SAR imagery, which contains details of the marginal ice zones, in tandem with sea ice concentration from passive microwave sensors, to obtain an improved estimate of sea ice concentration. To estimate ice concentration, we propose a deep learning model with a U-net architecture, which is in essence a neural network with convolutional layers and skip connections that lacks fully connected layers. This architecture has significant benefits over a traditional convolutional neural network (CNN), which is more widely used for remote sensing problems. The U-net is better designed for per pixel predictions because it aggregates information over large regions [40].

The key objective is to determine if the models we have developed can overcome problems of passive microwave data when predicting sea ice concentration in the marginal ice

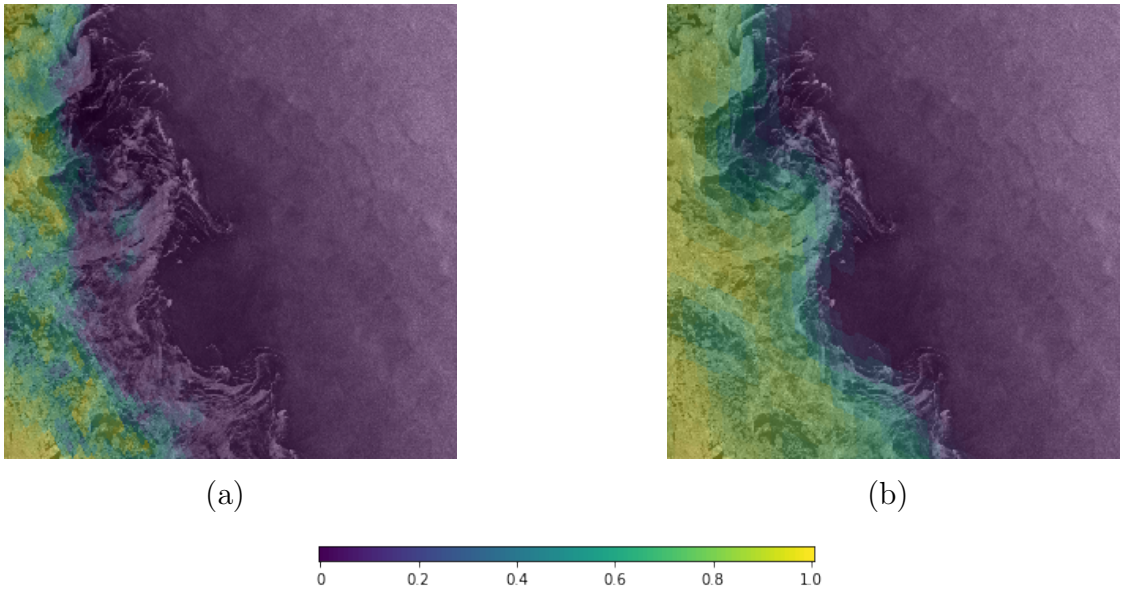


Figure 1.1: Passive microwave data based estimates of ice concentration overlaid on a marginal ice zone region of HH SAR image dated 02/27/2018. (a) The estimate, acquired from the ARTIST Sea Ice algorithm, predicts consolidated ice regions as lower ice concentration. Although some of this difference could be due to difference between the acquisition times of the passive microwave and SAR data, a bias of this nature is observed in studies of images along the Labrador Coast [64]. (b) The estimate, acquired from the NASA Team 2 algorithm, lacks details in the marginal ice zones.

zone while still maintaining good sea ice concentration predictions in regions of consolidated ice and open water. In this study, we train our model with passive microwave data where there is good visual agreement with SAR imagery. Therefore, our investigation is centered around the ability of our models to generalize information from this dataset to unseen images, which have visually incorrect labels in the marginal ice zones. Furthermore, we explore parameters that help us achieve this objective and show their effects. We also show the improvements a U-net provides over the traditionally used CNNs for this problem.

Chapter 2

Background

This chapter provides background information on concepts relating to this thesis. The first two sections of this chapter cover basic knowledge in sea ice and sea ice concentration required for this thesis. The following two sections provides information on the datasets used in this thesis. Here, the study area and study period are introduced as well as the different types of data used in this study. The next section is an overview on the deep learning concepts that are applied in this thesis. Finally, this chapter closes with previous work in this domain.

2.1 Sea Ice

Sea ice is frozen ocean water [19]. Ice that is floating in water can be categorized to five groups depending on their age and stage in development. They are summarized as follows:

1. New ice: Ice that has recently formed and composed of ice crystals weakly held together. This category of ice consists of frazil ice, grease ice, slush, and saga [19].

2. Nilas: Thin sheet of ice that is less than 10cm in thickness [51].

3. Young ice: Transition phase between nilas and first year ice that are 15-30cm in thickness. Can be subdivided into grey ice and grey-white ice. Grey ice is 10-15cm in thickness and can raft under pressure. Grey-white ice is of thickness 15-30cm and form ridges under pressure [19].

4. First year ice: Ice that is over 30cm in thickness and has not existed for more than one year [19].

5. Old ice: Ice that has survived for over one year and can be subdivided into second-year ice and multi-year ice [19].

Other than the category of sea ice, there are additional sea ice features that are important. Some of them are listed below:

1. Ice floes: Sheet of ice floating in water [51].

2. Ice eddies: Circular current of water and ice [48].

3. Rafting: Process by which thin ice slides over one another due to wind and ocean currents moving them around [51].

4. Pressure ridges: Ice blocks that form when ocean currents or wind exert force on ice floes making them converge and pile up [51].

Further information on sea ice can be found in other sources, such as the Manual of Ice (Government of Canada)[19] and a cryospheric glossary from the National Snow and Ice Data Center [51].

Information on sea ice has many practical applications. For example, pressure ridges are a hazard for ships and can immobilize them [44]. Sea ice thickness and extent are also climate indicators. Therefore, it is desirable to have a good understanding of the presence, movement, and mechanics of sea ice through time using a wealth of available remote sensing datasets.

2.2 Sea Ice Concentration

Sea ice concentration is a numeric value between 0 and 1. It is calculated for a unit area as the area of ice cover divided by the total area. An ice concentration value of 0 represents no ice cover in an area. An ice concentration value between 0 and 0.1 is referred to as open water [19]. An ice concentration value of 1 represents 100% ice cover in a region. This is referred to as consolidated ice [19]. Calculating ice concentration can help ships avoid consolidated ice regions and help scientists understand climate trends.

2.3 Study Area and Study Period

The study area, Hudson Strait, which connects the North Atlantic to Hudson Bay, was chosen for its current usage and anticipated usage by shipping vessels as well as its ice conditions [8]. The Hudson Strait is a route shipping companies use to access communities and mines along the coast of the Hudson Strait and Hudson Bay. Expected reduction in sea ice cover may make this route viable and more widely used in the future [57]. Vessels travelling through icy waters, as seen in Hudson Strait, would benefit from this study because improved ice concentration estimates at higher spatial resolution would help vessels navigate safely in this region [44, 5].

The study period is January - March 2018. This time period was chosen because it is a typical time period over which the majority of ship besetting incidents occur [44]. In this time period, a variety of ice types can be seen, such as new ice, ice floes, ice eddies and filaments, in addition to grey, grey-white, and thicker first year ice. Additional images from the Gulf of St. Lawrence (February 2011), Beaufort Sea (April 2015/2016), and Labrador Sea (February 2011) were also included in the training as a form of data augmentation to increase the number and variety of consolidated ice and open water conditions. Additional scenes from the High Arctic (July - September 2013) were used for evaluation. These regions are shown in Figure 2.1.

2.4 Standard Sea Ice Concentration Retrieval Methods

Satellite sensors are the preferred method to monitor sea ice for its ability to capture large spatial regions. Among satellite data, optical data is not preferred due to interference from cloud cover. Water droplets in clouds cause non-selective scattering of the visible electromagnetic radiation that is reflected by the Earth [49]. Passive microwave data and SAR data utilize microwave signals, which are longer wavelength than the visible spectrum and can penetrate clouds [49]. This makes SAR and passive microwave data the data source of choice for sea ice concentration retrieval. Optical data, such as Landsat data, can be used for verification purposes in regions without cloud cover. These data sources are described in this section.

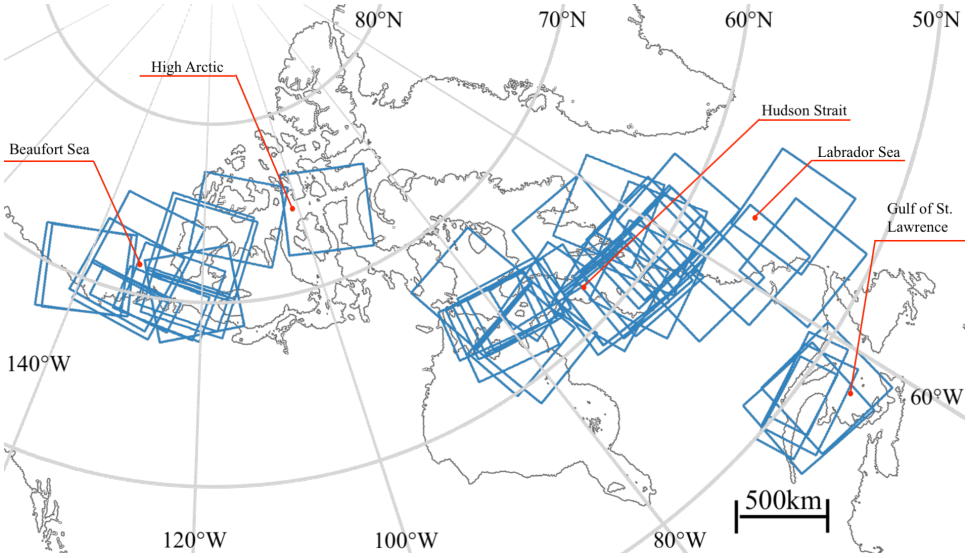


Figure 2.1: Scenes taken from Hudson Strait, Gulf of St. Lawrence, Labrador Sea, Beaufort Sea, and High Arctic were used in this study. Footprints of acquisitions used in this thesis are provided as hollow blue polygons. Canada coastline for this figure was acquired from the landmask of the ASI data [61].

2.4.1 Passive Microwave Data

Passive microwave data is the most popular data source for sea ice concentration information retrieval [47]. Passive microwave sensors measure the microwave radiation naturally emitted by the Earth. Clouds have very little effect on passive microwave data because they emit very little microwave radiation, compared to sea ice [52]. Physical properties of an object, such as atomic composition and crystalline structure, determines the emitted radiation [52]. Passive microwave data is advantageous to use for sea ice concentration estimation because there is a strong contrast in the emissivity of ice and water [10]. The limitation of passive microwave data for sea ice concentration estimation is that the kinetic temperature of the sea ice and open water are such that they emit radiation at a low rate, causing the need to collect data over large swaths of the Earth's surface to obtain an observation with the sensor onboard the satellite. Therefore, passive microwave data is characterized by low spatial resolution when compared with optical and SAR data. Consequently, sea ice concentration retrieval methods from passive microwave data are not able to capture fine details in sea ice, such as leads [52]. The ice concentration obtained also varies non-linearly with the size of the swath. Therefore, there can be disagreement

between estimates obtained from passive microwave data and estimates obtained from other retrieval methods. Among passive microwave data channels, the frequency of the microwave signal determines the spatial resolution of the data. Signals of higher frequency have finer resolution, but they suffer from more atmospheric contamination. Most passive microwave data based sea ice concentration retrieval algorithms use a weather filter, which is a bulk correction for atmospheric opacity that sets the ice concentration to zero. The weather filter is problematic because it is difficult to distinguish an atmospheric signature from intermediate ice concentration in the marginal ice zone and sometimes the weather filters remove thin or diffuse ice instead of erroneous retrievals due to weather [23]. There are many methods to obtain ice concentration estimates from passive microwave data [23, 2], estimates obtained from two algorithms are used in this thesis.

In this thesis, we use ice concentration estimates obtained from the Institute of Environmental Physics at the University of Bremen as training labels [61]. These estimates were obtained by applying the ARTIST Sea Ice (ASI) algorithm [61] to microwave radiometer data of the sensors AMSR-E (Advanced Microwave Scanning Radiometer for EOS) on the NASA satellite Aqua and AMSR-2 (Advanced Microwave Scanning Radiometer 2) on the JAXA satellite GCOM-W1 [61]. The 89GHz channels were used because they provided the finest spatial resolution among all channels from AMSR-E and AMSR-2. The distance between successive ice concentration estimates is 3.125km. The higher frequency channels used by the ASI algorithm may have fine resolution, but they also have some shortcomings. The 89GHz channels are heavily influenced by atmosphere conditions such as water vapour and cloud liquid water. This results in noisy ice concentration estimates. The weather filters used by this data source use lower frequency channels of the passive microwave sensors, which have coarser resolution and results in a bias towards low ice concentration values. Hereafter, we refer to ice concentration estimates obtained from the aforementioned source as ASI estimates.

For comparison, another ice concentration estimate obtained from the NASA Team 2 (NT2) algorithm [41] was used in this thesis. The NT2 algorithm is an enhanced version of the NASA Team (NT) algorithm [41] that estimates ice concentration from passive microwave data of frequencies 18.7Ghz - 89Ghz [39]. The distance between successive ice concentration estimates is 12.5km.The lower frequencies are affected less by atmospheric conditions and weather. Due to less atmospheric contamination, the effects of a weather filter is not as prevalent in NT2 data when compared with ASI data. Therefore, ice concentration estimates obtained from this source are less likely to have a low bias, but they cannot provide the spatial details that the 89GHz channels from the ASI data provide.

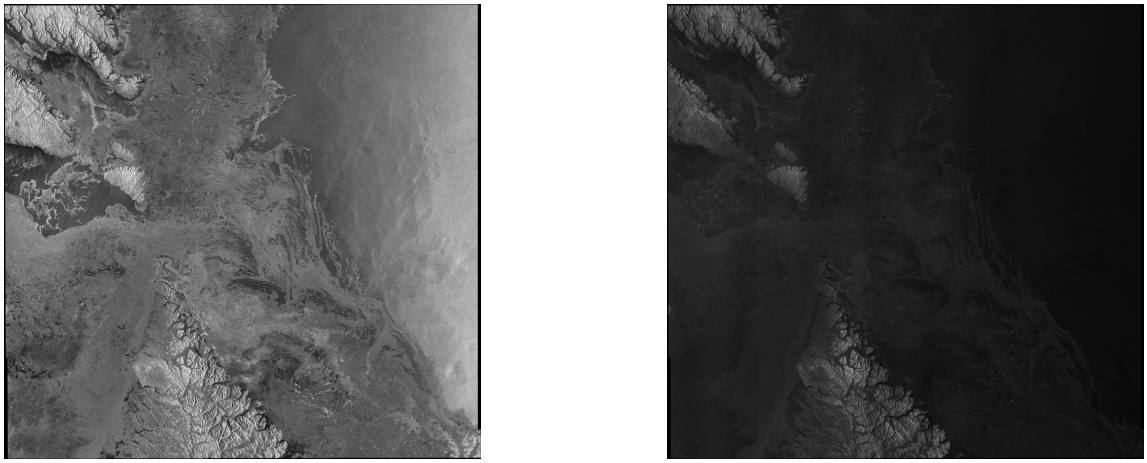
2.4.2 Synthetic Aperture Radar (SAR) Data

SAR data is another data source that can be used to obtain sea ice concentration information. It is obtained from multiple sensors that send a microwave signal to the Earth and records the signal reflected back [46]. This is beneficial over passive microwave data as there is control over the amplitude and frequency of the signal transmitted. Therefore, SAR data can be of higher spatial resolution than passive microwave data. The interaction of the SAR signal with the Earth is dependent on the wavelength of the signal. Generally, the SAR signal interacts most with items closest in size to its wavelength [24]. Therefore, different bands of SAR signals are suitable for different tasks. C-band SAR data, which is of wavelength 7.5-3.8cm, was found to be most suitable for ocean and ice monitoring [46] and is used in this thesis. SAR signals of this wavelength are not affected by cloud and atmospheric contamination as those particles are much smaller than the wavelength of the SAR signal. The backscatter of the SAR signal is dependent on the surface characteristics, such as structure and roughness, the incidence angle, and the polarization of the SAR signal [43]. For example, open water that is smooth appears dark as the transmitted signal is reflected away from spacecraft [24]. Rougher surfaces appear brighter as the transmitted signal is scattered in all directions. Images acquired from SAR data are referred to as SAR images.

Polarization is the orientation of an electromagnetic signal. SAR sensors are able to control the orientation of the signal in the transmit and the receive. There are four different orientations available for SAR images. They are listed below:

1. HH Polarization: Signal is oriented horizontally in transmit and horizontally in receive.
2. HV Polarization: Signal is oriented horizontally in transmit and vertically in receive.
3. VH Polarization: Signal is oriented vertically in transmit and horizontally in receive.
4. VV Polarization: Signal is oriented vertically in transmit and vertically in receive.

Since backscatter is dependent on polarization, using multiple polarizations of SAR signals provide different information on the surface of the Earth [43]. One study found that HV channels provide good floe delineation, but poor discrimination of new thin and smooth first year ice types [4]. They also observed various levels of backscatter in HH SAR images, which made it difficult to distinguish consolidated ice from open water in some regions [4]. Similar observations were made from the dataset used in this thesis. Therefore,



(a) (b)

Figure 2.2: SAR images of (a) HH and (b) HV polarization taken from the Hudson Strait on 01/20/2018. Images of HH and HV polarization look different and provide different information because they interact differently with the Earth's surface.

SAR data of HH and HV polarization are used in this thesis. HH and HV polarized SAR images are shown in Figure 2.2. HH polarization was used because it captures more ice and water features from the image, which can help the model determine ice concentration. HV polarization was used because it is less affected by varying incidence angles in comparison to HH polarization.

Using SAR scenes for ice concentration estimation is advantageous because they are becoming increasingly more available and they are of higher spatial resolution [53]. Images used in this thesis are derived from the RADARSAT-2 ScanSAR Wide mode. 8-bit images are used in this thesis. The distance between successive pixels on these SAR images is 50m.

Although there are many benefits to using SAR images for ice concentration estimation, there are also limitations. First, the interaction between the SAR signal and ice and water is very complex. The aformentioned observation that smooth water appears darker is not always the case as the incidence angle can make smooth open water appear bright in certain regions [4]. From both HH and HV SAR images shown in Figure 2.2, it can be seen that intensity alone cannot be used to distinguish ice from water and therefore does not contain sufficient information for ice concentration retrieval. Second, SAR images often suffers from image noise, such as speckle noise and banding effect. The banding effect can be seen

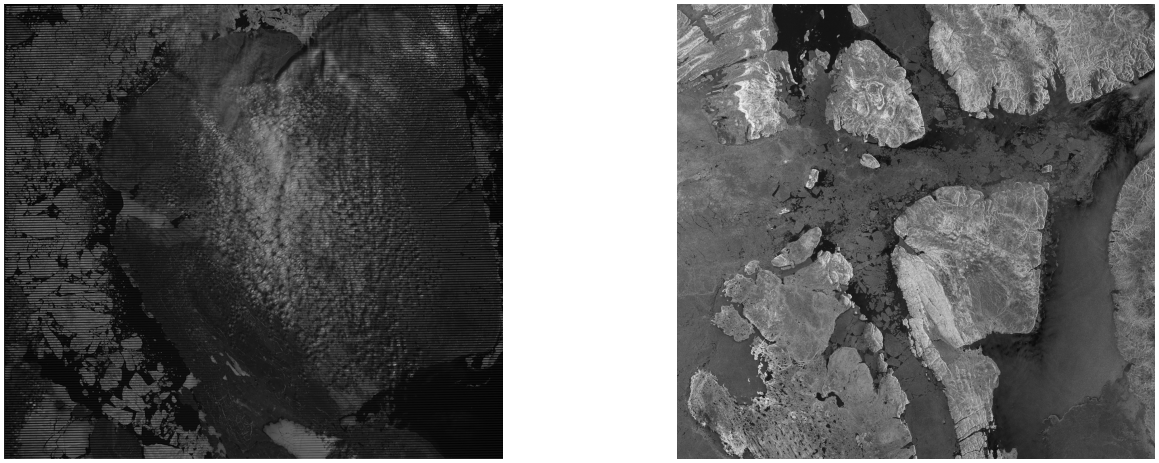
on the SAR image of HV polarization in Figure 2.2. Lastly, sea ice concentration retrieval from SAR images is not straightforward because there is very little information obtained from a specific pixel on a SAR image. Spatial information is required to determine the ice concentration.

Sea ice concentration can be determined from visual interpretation of SAR imagery or automated models. Image analysis charts are ice expert renditions of SAR imagery. Regions of homogeneous ice conditions (polygons) are assigned an ice concentration value in tenths (ex. 0.1, 0.2...). Therefore, image analysis charts are not as numerically precise as algorithms using passive microwave data. Furthermore, image analysis charts are subject to the interpretation of the expert doing the rendition. Studies have shown variation in assigned labels from different analysts for the same region [42]. With the increasing amount of SAR data becoming available, automated methods of ice concentration retrieval from SAR imagery are highly desirable. SAR images and image analysis charts used in this thesis were provided by Environment and Climate Change Canada (ECCC).

2.4.3 Landsat Data

Landsat images are high resolution satellite images from optical and near-infrared sensors. Optical and near-infrared signals have shorter wavelength than microwave signals used in SAR data and passive microwave data, which make data retrieved from these signals of higher spatial resolution. The limitation to using electromagnetic waves in the optical and near-infrared range is that they do not penetrate clouds [49]. This is a severe limitation because sea ice often occurs in regions of cloud cover [59].

For this thesis, band 8 from Landsat-7 images were used to retrieve sea ice concentration estimates from regions without cloud cover. Band 8 registers wavelengths in the range 520-900nm, which is the optical and near-infrared range, and it was chosen because it provides the highest spatial resolution, 15m, from all bands. As seen in Figure 2.3, Landsat images capture more details than SAR images. Albedo information, which is a measure of the incident electromagnetic radiation reflected by a surface [50], can be derived from Landsat images, which can be converted to ice concentration estimates with a threshold analysis [67]. The albedo value of a material to electromagnetic radiation is dependent on the wavelength of the electromagnetic radiation and the physical properties of the material. An albedo value of 0 means the surface is a perfect absorber, absorbs all incoming electromagnetic radiation, and an albedo value of 1 means the surface is a perfect reflector, reflects all incoming electromagnetic radiation. In most cases, open water has an albedo of below 0.1 and sea ice has an albedo of around 0.5-0.7 in the visible spectrum [50]. Outside this range,



(a) (b)

Figure 2.3: (a) Landsat image dated 07/31/2013 and (b) HH SAR image taken 07/31/2013 from the High Arctic. Landsat image has finer resolution and smaller details can be seen.

sea ice has been observed demonstrating an albedo below 0.1 more often [36, 56].Therefore, for the threshold analysis, an albedo threshold of 0.1 is used where all pixels below an albedo of 0.1 is set to 0% ice concentration and above is set to 100% ice concentration. The Gaussian weighted mean is then taken from the 10×10 Landsat pixel blocks to obtain an ice concentration label [67]. The pixel spacing of the Landsat ice concentration estimates was 150m, which is finer than all other sea ice concentration estimates used in this thesis.

Previous studies have found that the albedo of sea water to be within 0.01 and 0.05 for wavelengths of range 200-2500nm, with the highest albedo of 0.05 in the visible spectrum range [29]. Studies have also found that there is a variation in the albedo of sea ice depending on the type of ice (snow covered ice, grey ice, white ice etc.) [36, 56]. Generally, sea ice has an albedo of above 0.5 in the visible spectrum, but certain ice types, such as dark nilas, have been observed to have an albedo below 0.1 in the visible spectrum [56]. Therefore, the threshold analysis method with albedo in the optical and near-infrared range also has some limitations.

Hereafter, we refer to ice concentration estimates obtained from the Landsat data as Landsat estimates. Landsat-7 images are courtesy of the U.S Geological Survey.

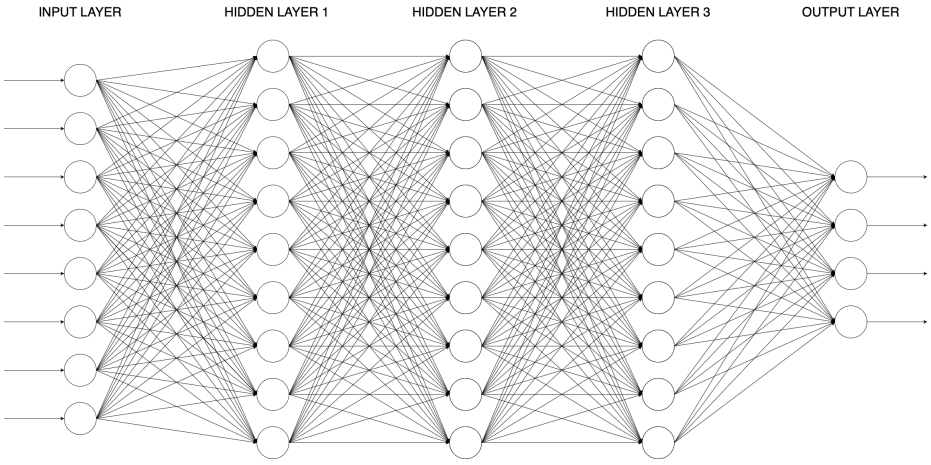


Figure 2.4: Example of a deep neural network consisting of an input layer, output layer, and three hidden layers. Each node performs mathematical computations with learnt weights and outputs a value.

2.5 Deep Learning

Deep learning is desirable for this study for its capability to interpret complex data types and for the fact that these methods are able to learn features from data, in contrast to methods [7, 32] that require features to be specified. Deep learning is a machine learning method where complicated concepts are learnt by building on simpler ones [18]. This thesis focuses on deep neural networks, which are multi-layer networks that are able to learn complex relationships. Deep neural networks consist of an input layer, which reads the initial data, an output layer, which outputs the desired values, and hidden layers, which are layers in between the input and output layers. An example of a deep neural network consisting of an input layer, an output layer, and three hidden layers is shown in Figure 2.4. Each layer in a deep neural network performs mathematical computations with weight matrices. Weight values are usually determined through training with large datasets. More information on deep learning can be found outside this thesis [18].

2.5.1 Deep Learning Model Design Choices

All deep learning models require user defined design choices. Some design choices crucial for this thesis is outlined in this section.

First, each layer in a neural network is followed by an activation function. The activation function is often a elementwise non-linear transformation that introduces non-linearity to the output. Sigmoid, rectified linear unit (ReLU), and Tanh functions are examples of activation functions. The models in this thesis use ReLU activation function for the hidden layers. ReLU is the preferred choice for activation functions because it has proven to lead to faster learning, and it does not saturate like other activation functions [45, 30].

Another design choice is the loss function (sometimes called cost function). The loss is the difference between model output and training labels. It is a metric to measure how close the prediction is to the training data using a specified norm. The goal of deep learning models is to minimize the loss. Therefore, it is important to select the loss function that measures the ability of the model to achieve the project goal. Mean squared error (L2) and mean absolute error (L1) are common loss functions for models performing regression. The loss functions are defined as follows,

$$Loss_{L1} = \frac{1}{n}\sum_{i=1}^{n} |f(x^{(i)}, \theta) - y^{(i)}| \tag{2.1}$$

$$Loss_{L2} = \frac{1}{n}\sum_{i=1}^{n} (f(x^{(i)}, \theta) - y^{(i)})^2. \tag{2.2}$$

In Equation 2.1 and Equation 2.2, the output from the model, f, with parameters, θ, is represented as $f(x^{(i)}; \theta)$. The number of samples to calculate the loss is represented as n. The training label value for sample i is represented as $y_{(i)}$.

L1 loss takes the absolute value of the difference between the training label and the prediction (L1 norm) whereas L2 loss takes the square of the difference (L2 norm). Note that because the L2 loss takes the square of the difference, large differences will be amplified. This will be revisited later in the thesis.

The last design choice important for models in this thesis is the optimization algorithm. The optimization algorithm adjusts the weights of the neural network models to minimize the loss. There is no correct way to choose the optimization algorithm for a problem and often it is determined empirically [18]. Stochastic gradient descent (SGD) and Adam [28] are two popular optimization algorithms. SGD is a gradient descent algorithm where loss function is minimized iteratively by moving in the direction of greatest loss reduction, with randomly selected training samples from the training set [18]. Adam, derived from adaptive moment estimation, is a recent algorithm that gained popularity for its fast convergence and easy configuration [18].

2.5.2 Convolutional Neural Network

A convolutional neural network (CNN) is a deep learning model that consists of both convolutional and fully connected layers. The output of a convolutional layer is the convolution operation,

$$y[m,n] = h[m,n] * x[m,n] = \sum_{j=-\infty}^{\infty} \sum_{i=-\infty}^{\infty} h[i,j] \cdot x[m-i, n-j], \tag{2.3}$$

between the input and a user-specified number of filters. In equation Equation 2.3, the 2-dimensional output, y, is the result of the convolution operation, represented by $*$, between the 2-dimensional kernel, h, and the 2-dimensional input x. This provides the model with spatial information from the images. The output from a convolutional layer consisting of a number of filters applied to a 2-dimensional image is a 3-dimensional feature map ($length \times width \times depth$). Subsequent convolutional layers maintain the dimensionality of the feature map. Before the fully connected layers, the output from the convolutional layers are flattened to a 1-dimensional vector. This is generally a lexicographical stacking of the rows or columns of the output from the convolutional layers. Between convolutional layers, there are often pooling operations to reduce the overall size of the input to the next layer. Pooling operations replace a 2-dimensional block of values with a single value, which reduces the size of the output [18]. For example, max pooling is an operation where the block of values is replaced by the maximum within the block, whereas average pooling replaces the block by the average taken over the block.

2.5.3 Fully Convolutional Network

A fully convolutional network (FCN) is a deep learning model that is similar to a CNN because it has convolutional layers, but differs from a CNN as it does not have fully connected layers at the end [33]. FCNs are primarily used for semantic segmentation where each pixel is given a class label, but its property of making predictions on patches in one pass make it useful for assigning an ice concentration label at each pixel. The proposed method is an FCN based on a U-net architecture. A U-net is an FCN with encoding and decoding stages along with skip connections that are desirable for this study [58].

The encoding stage is a downsampling phase where feature extraction is performed on the raw input and it is represented as a smaller feature map. Downsampling is often performed by pooling operations. The decoding stage is an upsampling phase where the

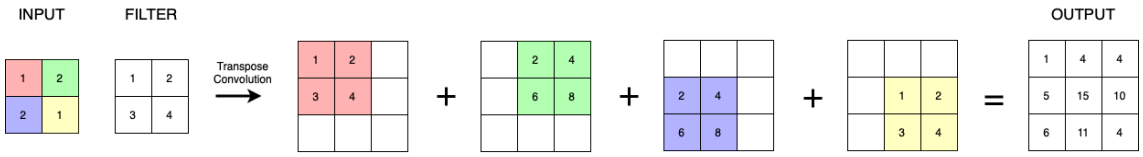


Figure 2.5: Example of a transpose convolution operation with a stride of 1 between a 2×2 input and 2×2 filter.

output from the encoding stage is processed to the original format. The upsampling operation is often done by interpolation, such as bilinear interpolation, or using learnable upsampling, such as transpose convolution. Transpose convolution is a convolution operation that differs from the traditional convolution operation, discussed in subsection 2.5.2, as it upsamples an image by multiplying each input element with a filter and mapping that to the output. Overlapping regions from adjacent pixel-kernel convolution operations are summed. An illustration of the transpose convolution operation is shown in Figure 2.5. Skip connections, which connect the downsampling and upsampling branches of the U-net, preserve finer details that may have been lost in the downsampling stages.

FCNs also differ from CNNs in the output they produce. The output of an FCN is a 2-dimensional matrix. An FCN can be trained to predict ice concentration values at each pixel location for the SAR image. This property allows the U-net to make predictions for each pixel location on the SAR image through one pass, whereas a CNN requires each pixel and its neighbors to be passed to make predictions. This results in a substantially faster run time for the U-net. The CNN is also limited by the receptive field of the patches. The U-net does not suffer from this problem because it can accept a full image as input and make predictions for each pixel (the SAR images were still split due to memory limitations in this thesis, but the region considered is still much larger than that for the CNN).

2.5.4 Curriculum Learning

Studies have found that models struggle to learn when presented with the full dataset at once [6]. When a model begins training with a subset of the full dataset and the remainder is progressively included, the model is able to obtain an understanding of groups sequentially. Curriculum learning is the structured training of a machine learning model by progressively increasing the difficulty of the training data [6]. This method has been proven to increase model performance, speed up convergence, and improve generalizability [66].

Curriculum learning is most commonly used for classification tasks where there is a disparity in the difficulty of the classes or in the difficulty within classes [34, 60, 6]. Curriculum learning has also been explored for classification tasks with class imbalance and shown success [65]. Applying curriculum learning for sea ice concentration estimation is desirable because marginal ice zones are more difficult to assign ice concentration values and it is not as prevalent as consolidated ice and open water regions.

2.6 Related Work

Labelling SAR images by their ice cover has been explored by supervised and unsupervised methods. Instance segmentation of sea ice images is the labelling of each pixel by its ice class. Region based segmentation, clustering based segmentation, and artificial neural network based segmentation are common methods to perform instance segmentation [27]. Region based segmentation is the grouping of image segments of homogeneous characteristics into regions and clustering based segmentation is the grouping of pixels of similar characteristics into clusters [27]. In the realm of SAR image segmentation, co-occurrence probabilities, Markov random fields, and Gabor filters have been used to extract texture features [13] for clustering and region based segmentation algorithms to discriminate ice from water on SAR images [15, 25, 71, 72]. After the rise of deep learning, traditional segmentation algorithms have not been preferred due to the use of rigid methods that require human intervention. With a greater number of SAR images and ice concentration labels being available, deep learning methods have garnered popularity for their ability to learn suitable features from SAR images.

Ice concentration estimation using automated models is the labelling of each pixel on a sea ice image with the ice concentration at that pixel. Deep learning for ice concentration estimation has been studied as early as 2014, where a multilayer perceptron neural network was trained using HH and HV SAR channels along with incidence angle information to obtain ice concentration estimates [26]. Since then, many advancements in deep learning have been made and deeper models have shown great success in ice concentration estimation. CNNs have been successful in estimating ice concentration on the Gulf of St. Lawrence when using image analysis chart estimates as training labels [64]. Convolutional layers allowed the model to learn spatial information, which enhanced performance in comparison to MLP (multi-layer perceptron) that does not take spatial information into account.

Using ice concentration estimates from passive microwave data to train a CNN model has also been studied. Passive microwave data has been used as training labels for a DenseNet model to estimate ice concentration in the Gulf of St. Lawrence and Arctic

Archipelago regions [14]. A DenseNet model is a CNN model that is unique in that there is a direct connection between all layers in a feed forward fashion [21]. This method reduces the vanishing gradient problem, allows features to be reused deeper in the model, and reduces the number of parameters. This thesis will be building on the aforementioned study. Furthermore, extensive studies on the shortcomings of passive microwave data for ice concentration estimation has also been explored [35, 23]. Estimates obtained from passive microwave data has been shown to be sensitive to the presence of melt water on the ice, atmospheric water vapour, and cloud liquid water [23, 2, 35].

U-nets have been used for a variety of segmentation problems across many domains [58], including remote sensing [74, 31]. Downsampling operations allow the U-net to capture long range dependencies and the upsampling operations allow the model to return an output of the same size as the input. Details that have been lost from downsampling operations are reintroduced in the upsampling layers through skip connections. These properties have been very useful for image segmentation, but they are also useful for regression problems. For example, per pixel regression with a U-net has been attempted for pansharpening applications and good results have been achieved [69]. They benefit from the encoding and decoding characteristics of a U-net, which is desirable for our study too. Another method to obtain long range dependencies without losing local details is using atrous convolution, which is a convolution operation where there are gaps between kernel weights [11, 12]. Atrous convolution is useful because it expands the receptive field to learn larger spatial features, but unlike downsampling, it does not lose spatial features. Previous studies have attempted this method for sea ice concentration and shown success [37].

Chapter 3

Methodology

In this chapter, the methodology for developing the proposed ice concentration retrieval model is discussed. This chapter consists of a data processing section to prepare the data for training and evaluation, a CNN architecture section, which describes the baseline CNN, a U-net architecture section that describes the proposed U-net, and finally an evaluation pipeline section that describes the procedure for evaluating the model performance.

In the following chapter, chapter 4, experiments are conducted to determine the best model. In chapter 5, the best model is chosen based on the results of the experiments. Subsequent chapters are evaluation of the proposed model.

3.1 Data Processing

In this section, processing of the data to train models is discussed. SAR images are processed to reduce noise and to be more manageable for training. Then estimates obtained from passive microwave data, image analysis charts, and Landsat data are used to annotate the images. After that, patches are extracted for model training and model evaluation.

3.1.1 SAR Image Processing

SAR images are very large files (approximately $10,000 \times 10,000$ pixel images) and they suffer from speckle noise. The distance between successive pixels on the SAR images was 50m. An 8×8 downsample was performed where 8×8 blocks were replaced with the

Table 3.1: Data used for training and evaluation.

Set	Location	Dates	No. Images
Basic Train	Hudson Strait	January-March 2018	22
Enhancement Set	Gulf of St. Lawrence	February 2011	6
	Labrador Sea	February 2011	4
	Beaufort Sea	April 2015	8
Hudson Strait Evaluation	Hudson Strait	January-March 2018	4
High Arctic Evaluation	High Arctic	July - September 2013	2

average value within the block. This put the distance between successive pixels as 400m and made the images more manageable for training and reduced speckle noise. To prepare the data for training and evaluation, the model input was created by stacking the HH and HV SAR images. We will refer to the 3-dimensional stacked images of shape $X \times Y \times 2$ as simply the SAR images.

From the full dataset, the data from the Hudson Strait was split into a training set, *Basic Train*, and an evaluation set, *Hudson Strait Evaluation*. Only scenes where there was a reasonable visual agreement between SAR and ASI estimates were used for *Basic Train*. Scenes that we wanted improvement on were reserved for *Hudson Strait Evaluation*. Another training set, *Enhancement Set*, was also made using scenes from Beaufort Sea, Gulf of St. Lawrence, and Labrador Sea. Another evaluation set, *High Arctic Evaluation*, was made using scenes from the High Arctic. The datasets are described in more detail in Table 3.1.

3.1.2 Image Annotation

The deep learning models require ice concentration values for each pixel on the SAR image for training and evaluation. Therefore, we annotate the SAR images with ASI, NT2,

Landsat, and image analysis chart estimates to have annotated images as training labels and evaluation labels. Since the SAR images have a different spatial resolution than the other data sources, we assign the label to each pixel in the SAR image using a nearest neighbour lookup process. The KDTree algorithm was used for this nearest neighbor lookup process [38]. Using these annotations, we generate 2-dimensional images of the same height and width as the SAR images. We use these annotated images as training labels and as labels to evaluate our model predictions.

3.1.3 Patch Extraction

Once the SAR images have been processed and annotated, the next step is extracting patches from the SAR images and the annotated images for training and evaluation. The CNN architecture requires patches for training and predictions while the U-net requires patches due to memory limitations. For the patches used for training, we extract them from regions away from land using a stride equal to half the patch size. The patch extraction procedure is slightly different for the CNN and U-net. For the CNN, we extract patches of sizes 25×25, 45×45, and 65×65. These patch sizes were used because visually they were found to provide varying levels of spatial information. Furthermore, a previous study found that the root mean squared error is lowest for patches of size 45 [63]. For the U-net, we extract patches of size 250×250. We only use the center 200×200 region to calculate the loss and update the weights. The region outside the center 200×200 region is padding, which ensures there is enough spatial information for the edge pixels on the center 200×200 region to make predictions. When making predictions, the CNN extracts patches with a stride of one and predicts ice concentration for the central pixel. The U-net extracts patches with a stride of 200 and predicts ice concentration for the central 200×200 pixel region.

3.2 CNN Architecture

The best CNN architecture for a problem depends on the problem objective and the data type. In this section, architectural choices for the baseline CNN are discussed.

The depth of a CNN determines the learning capabilities of a CNN. By increasing the depth, more downsampling layers can be added, which means larger spatial features can be learnt. Increasing the depth also means more parameters for the model to learn and therefore more learning capabilities. State-of-the-art CNNs are very deep (can be

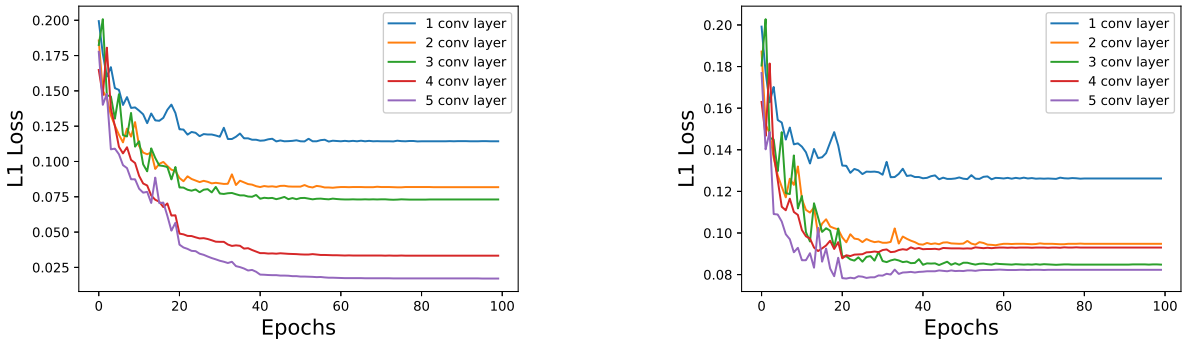


Figure 3.1: (a) Train and (b) test loss curves for CNN models of different sizes. Train and test data are from *Basic Train*. CNN models using 1 and 2 convolutional layers underfit the data as their train loss is higher than the rest. CNN models using 4 and 5 layers show overfitting as the test loss is much higher than the train loss. The model using 3 convolutional layers has very similar train and test losses. The CNN using 5 convolutional layers has the lowest test loss but the CNN using 3 convolutional layers was preferred for the fact that it does not overfit to the train set.

over 100 layers). The limitation of using deep networks is that they are more prone to overfit the training set. This is a possibility for sea ice concentration estimation because features are not as complex as those for which state-of-art CNNs are built. To determine the optimal number of convolutional layers for a CNN predicting sea ice concentration, the loss curves of five CNNs are compared, shown in Figure 3.1. For the comparison, the number of convolutional layers are varied while keeping the number of fully connected layers constant at two. All CNNs accept patches of size 45×45 and use an Adam optimizer. For the two largest models, bilinear interpolation was used to upsample the 45×45 patches to 200×200 patches to allow more downsampling layers. For this thesis, the CNN architecture chosen as the baseline consists of three convolutional layers and two fully connected layers. This architecture was chosen because underfitting is observed with more shallow models and overfitting is observed with deeper models, shown in Figure 3.1. Furthermore, when increasing the number of convolutional layers beyond three, the reduction in test loss is minor.

The input to the baseline CNN is a patch and the output is an ice concentration estimate for the center pixel of the patch. Each convolutional layer uses a 3×3 kernel and increases the depth of the feature map. Between each convolutional layer, there is a

2×2 max pooling operation, which replaces each 2×2 block with the maximum value in that block. This is a downsampling operation allows the 3×3 kernels to capture larger spatial information. The CNN architecture uses a ReLU activation function to introduce non-linearity between each layer [17]. The final fully connected layer was left to be linear as it was not followed by any activation functions. The CNN architecture used for this problem is shown in Figure 3.2. A larger model was not used because the features were not complex and increasing model complexity did not show extra value. The optimizer used for the CNN was Adam with an initial learning rate of 1×10^3 [28]. The Adam optimizer was selected after it obtained the lowest loss from empirical tests with a training and validation set. The learning rate was decayed gradually every 20 epochs by a factor of 10 for a total of 100 epochs. This learning structure was used because it took roughly 20 epochs for the validation loss to plateau at each learning rate and reduction in loss was minimal after 80 epochs.

Three different CNNs were made where each one accepts a different patch size. Different patch sizes were used for the CNN to study the effect of receptive field sizes. The three patch sizes, 25×25, 45×45, and 65×65, were selected based on their ability to capture different levels of spatial information. These patches are shown in Figure 3.3. From Figure 3.3a and Figure 3.3b, it can be seen that the small, 25×25, patch does not capture a lot of spatial information. Visually it is difficult to determine the ice concentration from this patch. Increasing the patch size introduces larger spatial features, as seen in Figure 3.3c and Figure 3.3f. At these sizes, it is clearer that the center pixel has an ice concentration value of 0. Although more spatial features are present at a patch size of 65×65, there are features that are not representative of low ice concentration within the patch. This can make it difficult for the model to predict the ice concentration accurately.

3.3 U-net Architecture

For this problem, a U-net consisting of 19 convolutional layers, 4 downsampling operations, and 4 upsampling operations was used. The U-net architecture is shown in Figure 3.4. Each convolution block, except the final block, in Figure 3.4 consists of two convolutional layers with ReLU activation [17]. The final convolution block consists of an additional convolutional layer that is not followed by an activation function. The first convolutional layer alters the depth when required, while the second layer maintains it. Convolution is performed with 3×3 filters for all convolutional layers. Downsampling was performed using 2×2 max pooling to obtain larger spatial information. Transpose convolution operation is used to upsample the image [73]. A stride value of 2 was used between each pixel

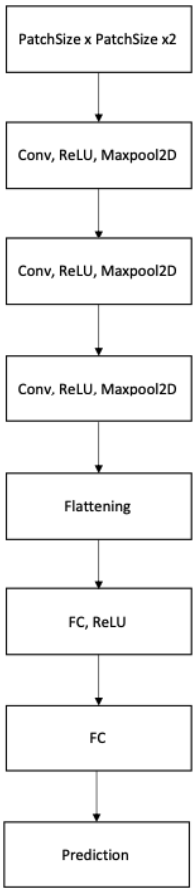
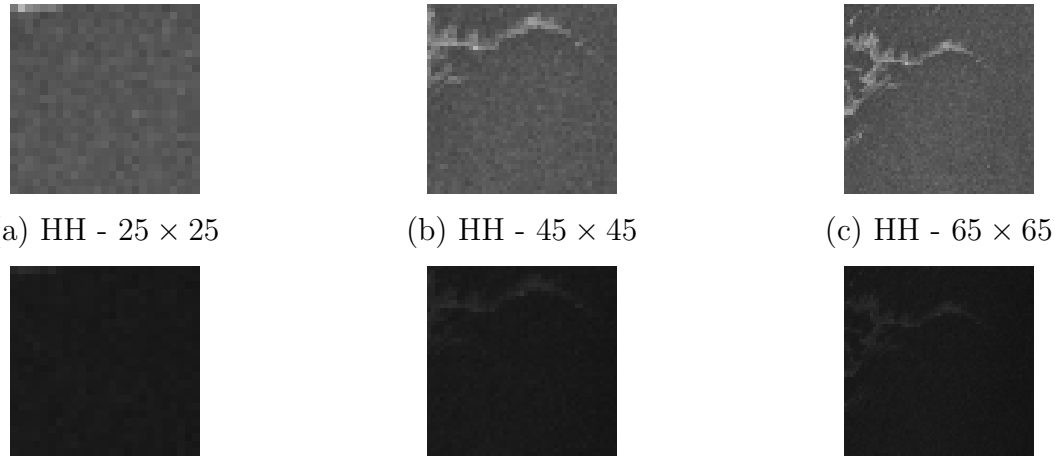


Figure 3.2: CNN architecture used in this thesis. Input size is dependent on patch size, which can be 25×25, 45×45, and 65×65.

operation. The skip connections between the downsampling and upsampling branches introduce smaller ice or water features that may have been lost in the downsampling branches. The optimizer used for the U-net was stochastic gradient descent with an initial learning rate of 1×10^1. SGD was found empirically to have the lowest training and validation loss. The learning rate was decayed gradually every 40 epochs by a factor of 10 for a total of 160 epochs. The decay rate was chosen to be 40 epochs as the validation and training loss plateaus at this point. Training was terminated after 160 epochs because after this point reduction in training and validation loss become negligible. The U-net was trained with patches of size 250×250 and they are shown in Figure 3.5.



(a) HH - 25 × 25 (b) HH - 45 × 45 (c) HH - 65 × 65

(d) HV - 25 × 25 (e) HV - 45 × 45 (f) HV - 65 × 65

Figure 3.3: HH and HV SAR image patches from SAR image taken on January 20, 2018. Larger patches capture more spatial information but features not relevant to the ice concentration of the center pixel are also picked up.

3.4 Evaluation Pipeline

The method to evaluate predictions is different for the CNN and the U-net. This results from the fundamental nature of both networks.

The CNN requires patches of the same size as the ones that were used to train it. The CNN also makes a prediction for a center pixel given the neighbouring spatial region. To make predictions on a SAR image, each pixel and its surrounding must be provided as input to the CNN. This is done by striding through the full image with a stride value of 1, extracting a patch around each pixel, and using that as input for the CNN. The downsampled SAR images are approximately 1200 × 1200 pixels. Therefore, the CNN is run approximately 1,440,000 times for each SAR image.

The U-net also requires patches due to memory limitations. The U-net differs from a CNN in the way it makes predictions for a full patch instead of just the center pixel. For predictions, 250 × 250 patches were extracted from the SAR image with a stride of 200. The center 200 × 200 predictions were only used. This was to remove any possible boundary effects. A 1200 × 1200 region can be split into 36 patches for which the U-net would make predictions. This is significantly less than the 1,440,000 predictions the CNN would have to make.

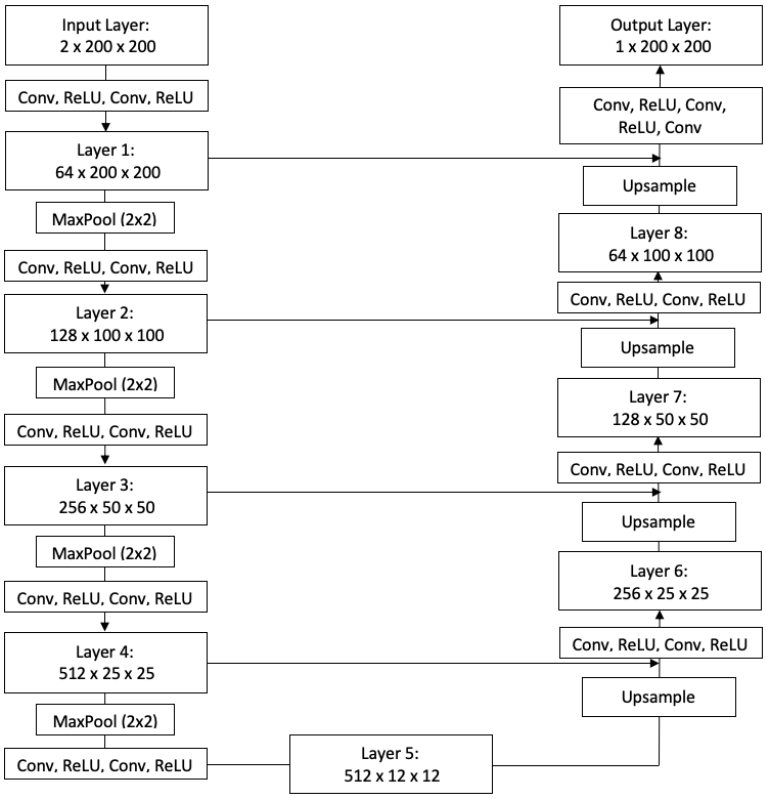
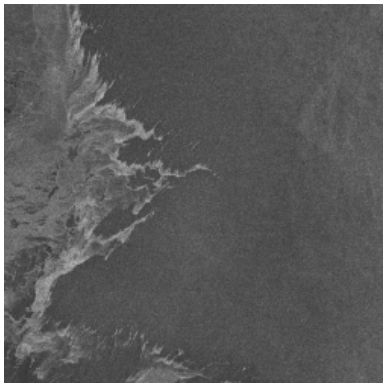


Figure 3.4: U-net architecture used in this thesis. Final convolutional layer is kept linear as it is not followed by any activation function.

Since the U-net and CNN have linear layers as the last layers, they make unbounded predictions. Since ice concentration values have a range of 0 to 1, the predictions were truncated to that range. Another technique to obtain a range of 0 to 1 is using a final activation function, such as sigmoid, that sets a range. It was preferred to keep the final layer linear and truncate the values rather than use an activation function that bounds the output because the latter was found to output images with noisier open water and consolidated ice regions. Activation functions output noisy open water and consolidated ice regions because when predictions are made, the model is more confident that certain regions are consolidated ice or open water than others. For example, prior to the activation function, pixels can have values of -0.2 or -0.3. Both values are strongly believed to be open water by the model, but when the final activation function assigns the range, it assigns different ice concentration values. In such a situation, it would be more ideal to give both pixels a value of 0.



(a) HH - 250 × 250

(b) HV - 250 × 250

Figure 3.5: 250 × 250 HH and HV SAR image patches from SAR image taken on January 20, 2018.

The last item of concern when making predictions is land and image boundaries. Land was ignored when making predictions. After predictions were made, a landmask was applied. CNNs are incapable of predicting at boundaries because it is not possible to get neighbouring region information for edge pixels. Zero padding and reflection padding are methods to allow CNNs to predict edge pixels, but they are not used here. The U-net makes predictions for a full patch of any size, so it is possible to make the SAR image edge as the edge of a patch. To make predictions from different models equal in size, regions that CNNs are incapable of predicting were removed. The CNN, which requires patches of size 65×65 has the largest region it cannot predict (32 pixels from every image boundary). Therefore, this region was discarded from all model predictions for comparison purposes.

3.5 Implementation

Python was used to build models required for this thesis. The PyTorch deep learning framework [54], which is a Python package, was used for the deep learning models. PyTorch provided easy implementation of typical deep learning functions (convolution, maxpooling, etc.) and performed back-propagation. Image annotations were completed using the scikit-learn package (sklearn) [55]. Sklearn provided a function to perform the KDTree nearest neighbour algorithm.

Chapter 4

Experiments

To choose the optimal model, the overall accuracy and the ability to generalize to many scenes was considered. The model must also overcome the low bias and coarse resolution issue, shown in Figure 1.1, of training with ASI estimates. It was found that a U-net using data augmentation, L1 loss, and trained with curriculum learning performs the best for this dataset.

To justify the decisions, 10 fold cross validation was performed with *Basic Train*. In this section, different loss functions, training datasets, and training sequences are compared. The models were also run on *Hudson Strait Evaluation*, shown in Figure 4.1, to visualize the effects of the design choices.

Figure 4.1a, Figure 4.1b, and Figure 4.1d were chosen because they contain large ice and water regions and ASI estimates often miss details in the ice edge between open water and consolidated ice. Furthermore, in Figure 4.1b, the bias towards lower ice concentration values of the ASI estimates are seen very clearly. The ASI algorithm predicts the ice concentration of consolidated ice regions in the vicinity of the ice edge as lower ice concentration. In Figure 4.1b, the ASI algorithm also predicts consolidated ice regions near land as lower ice concentration. Figure 4.1c was chosen because it contains smooth ice, which is ice that lacks typical features of ice, such as cracks. This can be problematic for neural network models predicting on SAR images as they can confuse smooth ice for water. Although the ASI algorithm does not struggle predicting the ice concentration for smooth ice regions, it struggles identifying cracks on the SAR image. The coarse resolution of the ASI estimates are unable to capture small cracks and precise boundaries for the cracks.

In this section, only the experiments, their results, and observations are shown. Discussion on results is left for chapter 5.

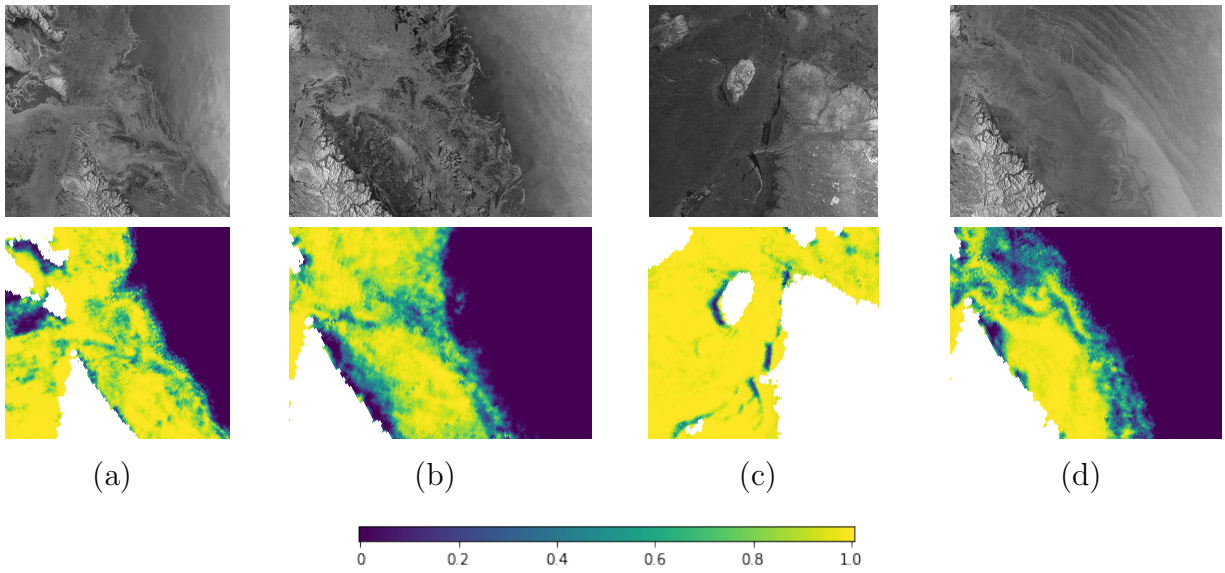


Figure 4.1: SAR HH polarization images and ASI estimates taken from Hudson Strait on dates (a) 01/21/2018, (b) 02/27/2018, (c) 03/04/2018, and (d) 02/03/2018. Details in the SAR images are being missed by ASI estimates. The marginal ice zone in (b) is very inaccurately predicted by ASI. The ASI algorithm also predicts ice near land as water in (b).

4.1 Effects of Receptive Field on CNN Models

CNNs with per pixel objectives use information from a spatial region to make a single prediction at a pixel that is usually located in the center of the spatial region. For the CNN, the input to the model is a patch and the output is the ice concentration for the center pixel. The model's receptive field is the size of the patch. Increasing the size of the patch would give the model more spatial information.

Three models that accept patch sizes of 25×25, 45×45, or 65×65 were trained to see the effects of receptive field on the model performance. The proposed U-net was also tested to compare the results with the CNN. The predictions from the CNNs and the U-net are shown in Figure 4.2. It can be seen that predictions made from the CNN trained with a patch size of 25 captures a lot of details in the SAR image but produces a noisy output. Predictions from the CNNs trained with patch sizes of 45 and 65 retains fewer details but is also less noisy. The U-net predictions, as seen in Figure 4.2m,n,o,p, does not have this problem and was preferred over traditional CNNs.

4.2 Loss Function

Previous studies in sea ice concentration retrieval using deep learning have used mean squared error (L2) and mean absolute error (L1) as the loss function [64, 14]. Choosing the appropriate loss function is important because the model will be optimized to minimize the loss.

The goal of the models is to estimate ice concentration accurately while being robust to incorrect values in the passive microwave training labels. L1 and L2 loss functions are commonly used for regression problems because they measure a numeric difference between predictions and labels. L2 loss function amplifies large errors while attenuating small errors. Although this is desirable in many cases as it is better to remove large errors before focusing on smaller ones, this is not desirable when there are incorrect values in the training labels. L1 loss function does not have these properties. In this section, the effects of training with L1 and L2 loss functions are explored to decide which one works better for sea ice concentration estimation.

Predictions made by U-nets using L2 loss function and L1 loss function are shown in Figure 4.3a,b,c,d and Figure 4.3e,f,g,h respectively. The model trained with L2 loss function has a low bias in all images and suffers from banding effect to a greater extent. This will be discussed in chapter 5.

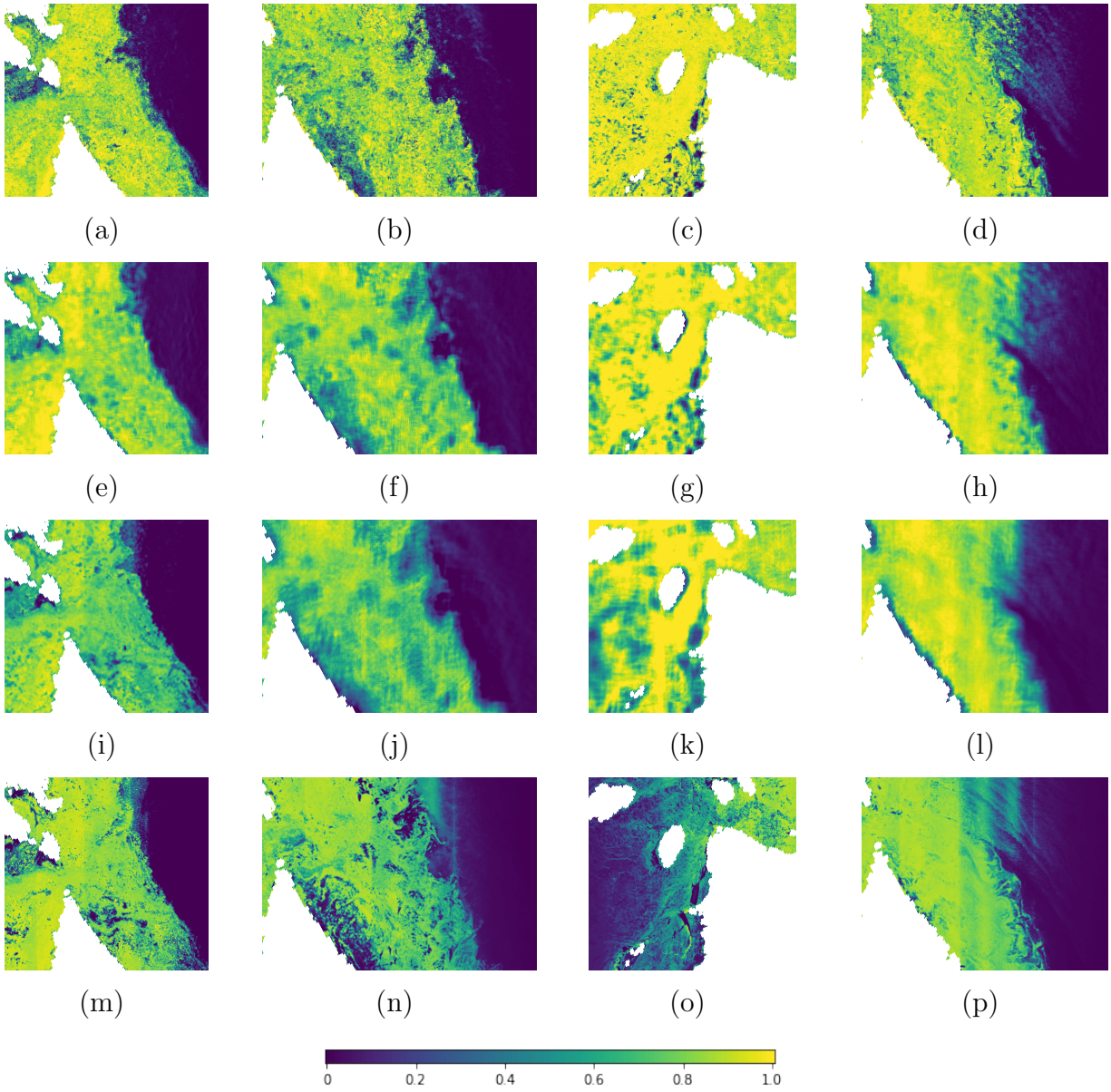


Figure 4.2: Comparison of CNN predictions using different patch sizes as input and of base U-net predictions. Predictions from model trained with patch size of 25×25 (a,b,c,d) is noisier but also captures more details from SAR images. Predictions from models trained with patch sizes of 45×45 (e,f,g,h) and 65 (i,j,k,l) are less noisy but also captures fewer details. U-net (m,n,o,p) does not have a receptive field issue.

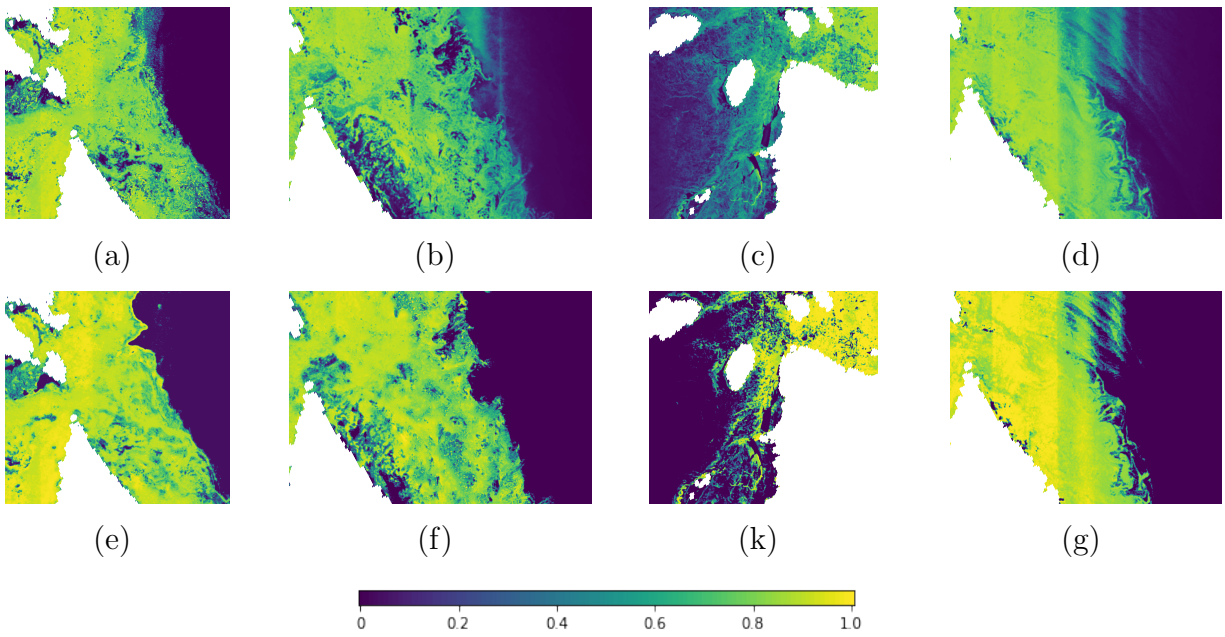


Figure 4.3: Comparison of the predictions of U-nets using different loss functions. Predictions made by U-net trained with L2 loss function (a,b,c,d) and U-net trained with L1 loss function (e,f,g,h) are shown. Predictions made by the model employing L2 loss function has a stronger bias towards lower ice concentration values.

4.3 Dataset Augmentation

To reduce the effect of incorrect values in the training data and to provide the models with more exposure, data augmentation was used. From our image scenes, selective sampling was used where only images that appeared accurate visually were used. From the 56 image scenes acquired from the Hudson Strait, 22 image scenes for which ASI ice concentration estimates appeared most accurate visually were used to train the models. Next, *Enhancement Set*, consisting of images from the Gulf of St. Lawrence, Beaufort Sea, and Labrador Sea, was added to the training set to provide more unique situations to the model. Although these images were taken at a different time period and different location, they provide valuable information on features present in different ice concentration levels. Lastly, during training, traditional data augmentation methods such as random rotation and reflection were incorporated to increase variation. Outputs from the U-net trained with *Basic Train + Enhancement Set* and L1 loss function is shown in Figure 4.4a,b,c,d.

Comparing with the models trained with *Basic Train*, it can be seen that the bias is improved. The consolidated ice region near land in Figure 4.1b is predicted correctly as high ice concentration. The banding effect issue is also eliminated. The model also performs better with smooth ice, seen in Figure 4.4g. Although increasing the data size improves the smooth ice, low bias, and banding effect issues, the amount of detail the model captures is also reduced. The model does not capture many details present in the marginal ice zones that the previous models captured.

4.4 Curriculum Learning

Curriculum learning experiments were conducted because, through experimentation, it was found that the models do not learn basic ice and water features when trained with marginal ice zones. When training with patches with consolidated ice and open water only, it was found that the model segregates ice and water very accurately, but it does not predict the marginal ice zones correctly. When marginal ice zones are included in training, there is a reduction in the sharpness of the ice edge. Moreover, for ice concentration estimation, it is more difficult for a model to estimate marginal ice than open water and consolidated ice.

Therefore, a curriculum learning approach was devised to ensure the model attains a strong understanding of consolidated ice and open water before incorporating samples representative of marginal ice zones. To do this, the training dataset was split into two groups, high ice concentration and low ice concentration set, based on average ice concentration

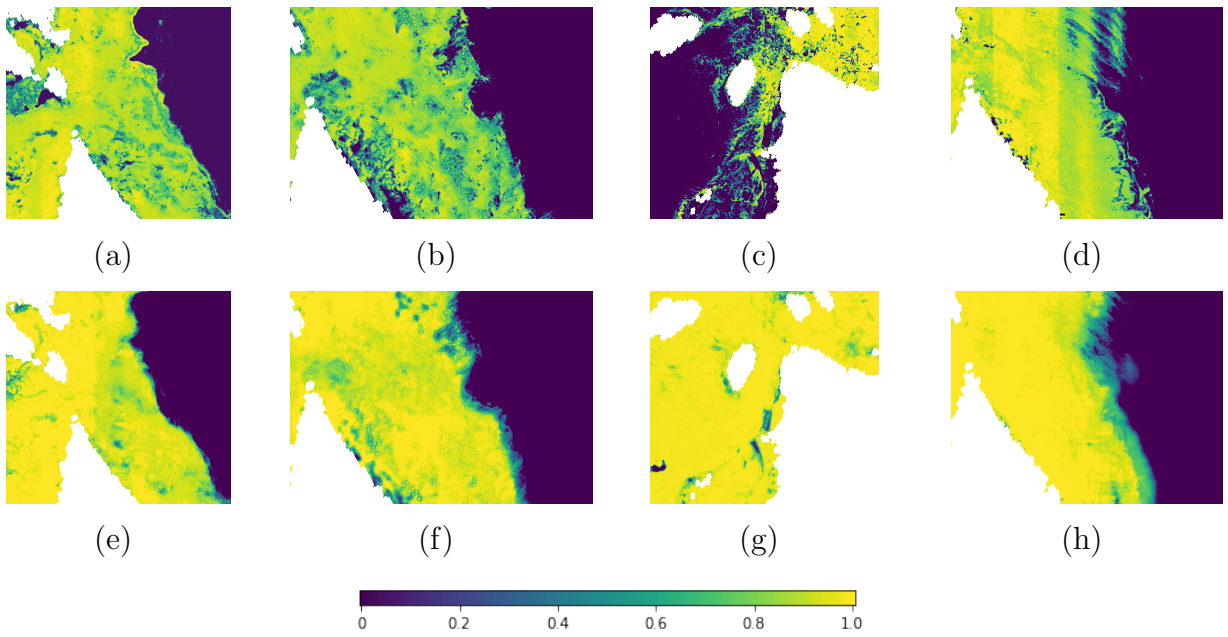


Figure 4.4: Comparisons of predictions from U-nets using different training datasets. U-net trained with L1 loss function (a,b,c,d) and U-net trained with L1 loss function with an enhanced dataset are shown (e,f,g,h). Enhancing the dataset reduces the low bias issue further but details from the SAR imagery are being missed.

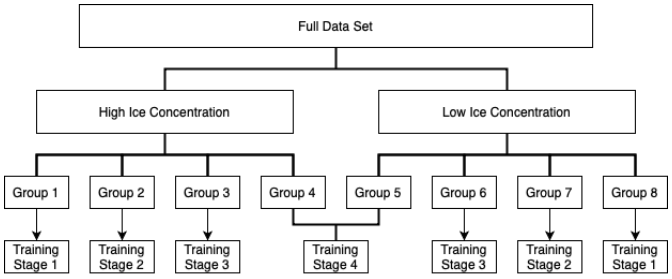


Figure 4.5: Curriculum learning procedure

in the patch. Next, four groups were formed within the two groups based on average ice concentration. Finally, opposing groups from the high ice concentration set and low ice concentration set were joined to form the four training stages. These training stages were numbered from one to four by increasing levels of difficulty (increasing marginal ice). This process is outlined in Figure 4.5.

The U-net, using enhanced data set with L1 loss function, was trained for a total of 140 epochs with an initial learning rate of 1×10^{-1}. Initially, the model was trained only with training stage 1. The other training stages were progressively added every 20 epochs. After training with the full set for 20 epochs, learning rate decay begins with a factor of 10 every 20 epochs to reach a final learning rate of 1×10^{-4}.

Results adding curriculum learning to the proposed model are shown in Figure 4.6e,f,g,h. The model captures more details that were originally missed while maintaining the benefits of the enhanced dataset. A shortcoming of applying curriculum learning can be seen in Figure 4.6f. Regions in the ice cover that are heterogeneous in appearance are being predicted as lower ice concentrations. This is not seen in the predictions made by the U-net using the enhanced dataset without curriculum learning (Figure 4.6b).

4.5 Choice of Training Labels

In previous experiments, models used ASI estimates as training labels. ASI estimates were preferred as training labels because they provide fine spatial resolution when compared with other passive microwave data based sea ice concentration retrieval algorithms. The limitation of ASI estimates for our study region is that they have lower ice concentration estimates in the marginal ice zone. In this section, training with NT2 estimates is explored in two methods. In the first experiment, NT2 estimates alone are used as training labels.

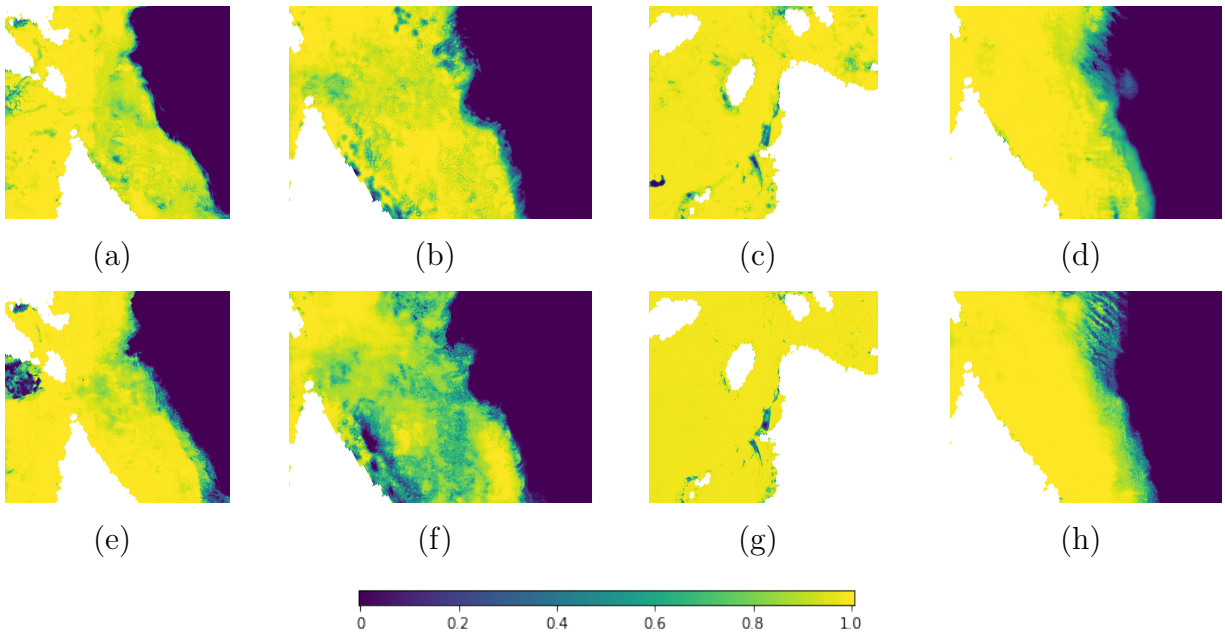


Figure 4.6: Comparisons of predictions from U-nets employing and not employing curriculum learning. U-net trained with L1 loss function with an enhanced dataset (a,b,c,d) and U-net trained with L1 loss function with an enhanced dataset using curriculum learning (e,f,g,h) are shown. The addition of curriculum learning retains more details while preserving the reduction of the low bias issue from the enhanced set.

In the second experiment, NT2 estimates are used alongside ASI estimates as training labels. For these experiments, the U-net, shown in Figure 3.4, is used with an L1 loss function and SGD optimizer. The training set used was *Basic Train* and curriculum learning was not performed. *Enhancement Set* was not used because the earliest year for which NT2 estimates are available is 2012 and *Enhancement Set* contains data from prior to that. Curriculum learning was not used because the goal of this experiment is to test model performance when training with NT2 estimates. Curriculum learning is sensitive to hyperparameter choices and the curriculum learning method was specifically chosen to overcome limitations of ASI data.

Three models were trained using different training schemes. The first model was only trained with NT2 estimates. The purpose of this experiment is to see differences when training with NT2 estimates. This model was trained for 160 epochs with an initial learning rate of 1×10^{-1} and a decay rate of 10 every 40 epochs. This was found to be the hyperparameters that resulted in the lowest training and validation loss. This model will be referred to as *Model-NT2* for simplicity. The second and third models used both ASI and NT2 estimates as training data and followed a two-phase training scheme. The second model was trained with ASI estimates before being trained with NT2 estimates. The third model was trained with NT2 estimates before being trained with ASI estimates. Each training phase for the second and third models was 160 epochs with an initial learning rate of 1×10^{-1} and a decay rate of 10 every 40 epochs. These two models will be referred to as *Model-ASI-NT2* and *Model-NT2-ASI* respectively.

The three aforementioned models were evaluated on *Hudson Strait Evaluation*. Results from this experiment are shown in Figure 4.7. Predictions from the U-net using ASI estimates as training labels, trained with *Basic Train* and L1 loss function is shown for comparison purposes. This model will be referred to as *Model-ASI* for simplicity.

4.6 10 Fold Cross Validation

10 fold cross validation is the primary method to justify the model choice. In this procedure, a training set, in this case *Basic Train*, is split into ten groups. One group is randomly chosen to be the validation set and the remainder of the groups are used for training. This process is repeated ten times and for each iteration, the loss is calculated. The mean and standard deviation among the training and validation loss values for the ten iterations are calculated. This information is used to determine the model that produces the lowest training and validation loss, the model that has the most stable training, and if any overfitting occurs.

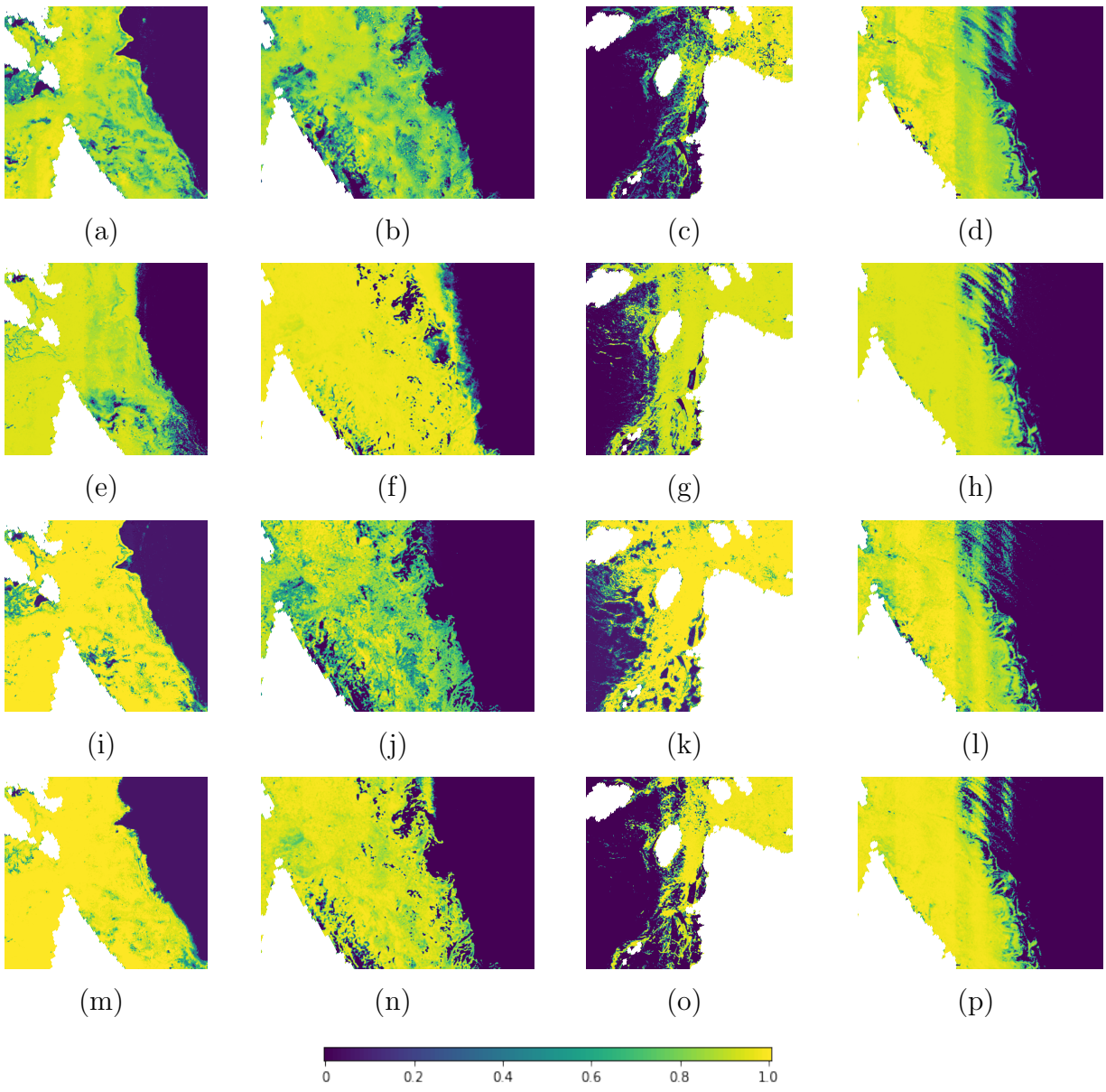


Figure 4.7: Comparison of U-net predictions made by *Model-ASI* (a, b, c, d), *Model-NT2* (e, f, g, h), *Model-ASI-NT2*(i, j, k, l), and *Model-NT2-ASI* (m, n, o, p).

Table 4.1: Mean and standard deviation of mean absolute error when performing 10 fold cross validation using *Basic Train*

Model	Train Mean (Std. Dev)	Test Mean (Std. Dev)
CNN-45	0.0787 (0.00367)	0.102(0.00376)
U-net (HS/L2)	0.136 (0.0116)	0.138 (0.00996)
U-net (HS/L1)	0.104 (0.00171)	0.108 (0.00161)
U-net (HS+E/L1)	0.0504 (0.0154)	0.0514 (0.0159)
U-net (HS+E/L1/Curr. Learning)	0.0385 (0.00397)	0.0387 (0.00387)

For this experiment, only certain models were compared. The CNN using 45×45 patches was selected from the CNN models because it had output of less noise than the CNN using 25×25 patches and more details than the CNN using 65×65 patches. All U-net models from experiments section 4.2-section 4.4 were used to evaluate the benefits of the proposed methods. Models using NT2 training labels from section 4.5 were not used because they could not provide information on proposed methods, such as curriculum learning and data augmentation, and they were not qualitatively superior to the U-nets using the *Enhancement Set*.

Note that the plots shown in this section and future sections are not the result of a model from the 10 fold cross validation experiment. They are all trained with the full *Basic Train*. 10 fold cross validation results are shown in Table 4.1.

Chapter 5

Analysis of Experimental Results

In this chapter, an analysis of the results and observations from chapter 4 is done. This chapter is divided into five sections, where the first section is overall observations from all experiments and the next four correspond specifically to the results and observations from the four experiments performed in chapter 4.

5.1 Overall Observations

When comparing model outputs, seen in Figure 4.1, Figure 4.2, Figure 4.3, Figure 4.4, and Figure 4.6, some clear differences can be seen. The deep learning models generally perform better at the ice edge and they do not predict consolidated ice near land as lower ice concentration to the same extent as ASI estimates. All models also predict water free from noise. This is partially due to the truncation of values to a range between zero and one. The models suffer with very rough water conditions shown in Figure 4.1d. The effects of rough water are prevalent in the SAR images and the models struggle to obtain the distinction between ice and water in that region. Patterns in the rough water are labeled as higher ice concentration. It must also be noted that rough water conditions were very limited in the training set. More exposure to such conditions may improve the performance of the models in this situation. It can also be seen in Table 4.1 that the average error on the training set is lower than the average error on the test set for all models. The CNN has a greater disparity between the train and test losses. This suggests that overfitting of the training set occurs. From Table 4.1, it can also be seen that the CNN has a lower training and test loss than the U-nets that do not use data augmentation. This is due to really poor predictions from the U-net on smooth ice regions.

5.2 Receptive Field Experiment

From Figure 4.2, the effects of the receptive field on CNNs can be seen. The CNN trained with a patch size of 25×25 output noisier images but with smaller details being preserved. Training with a small patch size resulted in noisier predictions because, as seen in Figure 3.3, it is difficult to determine the ice concentration of the center pixel when the patch is small. The CNN trained with a patch size of 65×65 made blurrier predictions that appear less noisy. Predictions made from models using larger patches are less noisy because the larger patch provides more spatial information. Therefore, a better understanding of the ice concentration of the center pixel can be made. Although a larger patch provides more spatial information, increasing patch size also has some limitations. When a large patch is used, spatial features unrelated to the center pixel may get captured in the patch and the model may get confused when determining the ice concentration for the center pixel. Another limitation for using a large patch is that patches for adjacent pixels will have a lot of overlap. Fine details, such as thin cracks in ice, will be missed because the 65×65 region surrounding a crack will be very similar to the 65×65 region surrounding an adjacent pixel that is not part of the ice crack. The deep learning model would struggle to differentiate the two patches due to the overlap. This results in blurrier outputs seen in Figure 4.2. The CNNs trained with patch sizes of 45×45 and 65×65 have low bias, which is not as prevalent in the CNN trained with a patch size of 25×25. This is also a result of the model having stronger predictions for ice and water when the patch size is 25×25. The U-net, shown in Figure 4.2, assigns labels to all pixels in one pass. Therefore, it does not have the receptive field issue where there is a trade-off between precision and noise.

5.3 Loss Function Experiment

Figure 4.3 shows the effects of using L1 and L2 loss functions during training. The predictions from the model trained with the L2 loss function retained more of the shortcomings of the ASI estimates. This occurs because L2 loss takes the square of the difference between the predicted value and the label field. Therefore, larger differences affect the loss terms more than small ones. Larger error values are often due to the ASI algorithm labelling ice regions as water. The L2 loss function forces the model to fit these situations to obtain a lower loss in the training set. Consequently, attempting to fit consolidated ice regions as water will make the model acquire a lower bias. Models trained with an L1 loss function are not affected to the same degree because there is no amplification of larger errors.

5.4 Data Augmentation Experiment

After data augmentation, shown in Figure 4.4, it can be seen that the model trained with the enhanced dataset had significant improvements. The low bias issue was reduced further because the model was exposed to different types and tones of ice and water regions. This allowed the model to generalize and learn high level features of ice without relying on intensity as much. Furthermore, the model trained solely with *Basic Train* performed poorly in smooth ice conditions. Smooth ice was problematic for models because these regions do not have one of the main common features of consolidated ice, which are cracks. Therefore, the lack of ice features can make the consolidated ice region appear like open water to the model. Additional ice images present in the *Enhancement Set* exposed the model to more instances of smooth ice. From Table 4.1, it can also be seen that data augmentation has significant improvements. The training and test mean error is lower for the model employing data augmentation.

5.5 Curriculum Learning Experiment

Lastly, from Figure 4.6, we can also see that curriculum learning behaves as we expect it to. The model using curriculum learning identifies details that were missed by the model trained with the enhanced set. This was the direct effect of the model obtaining an understanding of ice and water early on. We can also see from Table 4.1 that curriculum learning reduces the average train and test loss. The lower standard deviation also suggests that curriculum learning provides more stable training. In Figure 5.1, we overlay predictions from three models from a marginal ice zone region with the corresponding SAR image. We can see that the U-net trained without data augmentation captures the boundary between ice and water relatively well, but it predicts consolidated ice and marginal ice regions as lower ice concentration. The U-net using data augmentation does not have the bias issue, but it can be seen that details in the marginal ice zones are being missed. From the U-net using data augmentation and curriculum learning, it can be seen that curriculum learning solves the bias issue while preserving some of the details in the marginal ice zone. Furthermore, when applying curriculum learning, fewer epochs are required and early epochs are shorter because there is less data to train. Therefore, from this experiment, we show that curriculum learning achieves better performance in less time.

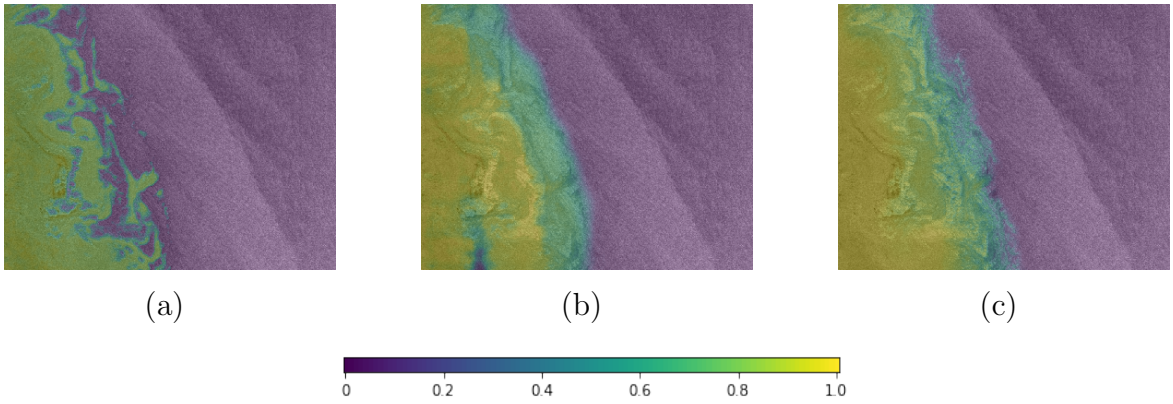


Figure 5.1: Predictions from marginal ice zone region from 02/03/2018 overlayed on HH SAR image. (a) U-net using L1 loss function (b) U-net using L1 loss function and data augmentation (c) U-net using L1 loss function, data augmentation, and curriculum learning.

5.6 Choice of Training Labels Experiment

From Figure 4.7, it can be seen that predictions made from *Model-ASI* has a strong bias towards low ice concentration as regions of consolidated ice are being predicted as lower ice concentration. Predictions from *Model-NT2* do not have a strong bias. Consolidated ice regions are generally being predicted as high ice concentration. The effects of training with coarse resolution training labels can also be seen in the predictions made by *Model-NT2*. This can be seen in Figure 4.7e, where the model fails to identify water near land and fails to capture the details in the marginal ice zone. Banding effect can be seen on all images, but it is more prevalent in the predictions made by *Model-ASI*. This can be seen clearly in Figure 4.7d. The models employing staged training suffer less from the banding effect and have less bias than *Model-ASI* and capture finer details than *Model-NT2*. It can also be seen that *Model-NT2-ASI* has a stronger bias towards low ice concentrations than *Model-ASI-NT2*. *Model-ASI-NT2* does not capture spatial details to the same extent as *Model-NT2-ASI*. From this observation, it appears that that the training labels the models were trained with last has the strongest effect on model performance. All model predictions perform poorly in the smooth ice region. This is expected because *Basic Train* does not have enough exposure to smooth ice.

5.7 Proposed Model

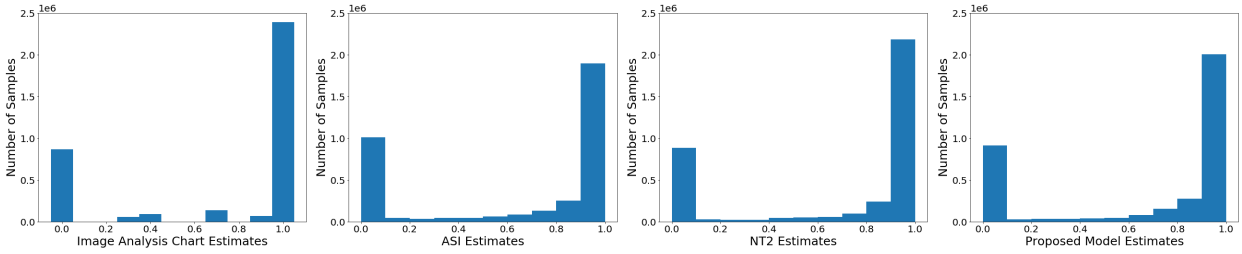
The following chapters consists of evaluation of the proposed model. Based on the experiments conducted in chapter 4, the proposed model was determined to be a U-net employing L1 loss function and trained with the enhanced dataset (*Basic Train + Enhancement Set*) while employing curriculum learning. This was decided to be the proposed model based on results from the 10 fold cross validation experiment.

Chapter 6

Hudson Strait Evaluation

In this section, the proposed model, U-net trained with the enhanced dataset using L1 loss function while employing curriculum learning, is evaluated on images from *Hudson Strait Evaluation*. The proposed model was evaluated on three unseen images for which we acquired image analysis charts. The specific dates from *Hudson Strait Evaluation* used were 01/21/2018, 02/27/2018, and 03/04/2018. Histograms breaking down the samples used for evaluation by ice concentration estimate value from different ice concentration retrieval methods is shown in Figure 6.1. It can be seen that the dataset is imbalanced and there are significantly more pixels labeled to have an ice concentration value between 0-0.1 and 0.9-1.0 than other values. This is representative of all image scenes because it is more common to have consolidated ice regions and open water than marginal ice zones.

It must be noted that there is no 100% accurate way to retrieve ice concentration. In



(a) (b) (c) (d)

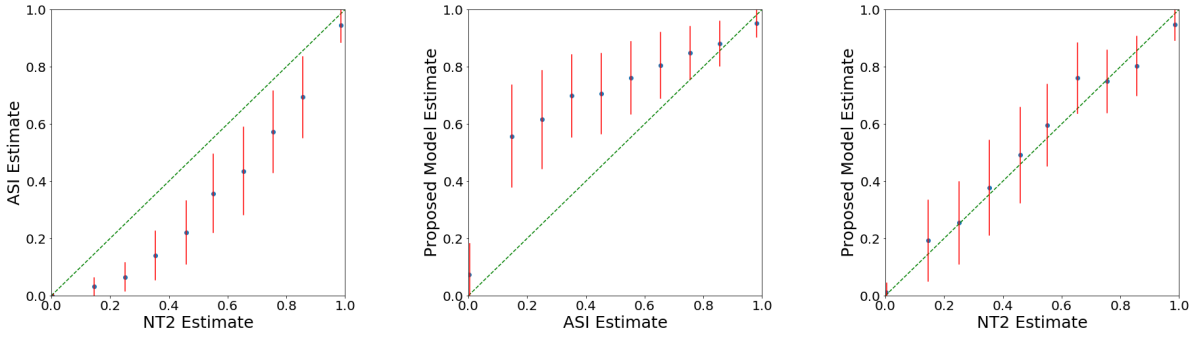
Figure 6.1: Histograms of ice concentration estimates from (a) image analysis charts (b) ASI (c) NT2 and (d) proposed model.

this chapter, a comparison is performed with estimates obtained from passive microwave data and image analysis charts. From this, an understanding of the relative performance of the different methods can be acquired. A qualitative comparison is also made to evaluate the ice concentration retrieval methods.

6.1 Passive Microwave Evaluation

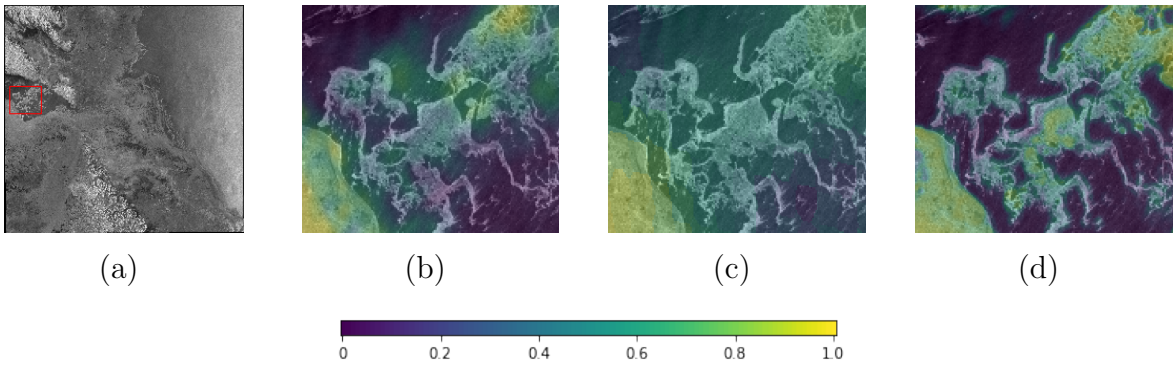
In this section, predictions from the proposed model are compared with estimates obtained from passive microwave data. To do this, estimates from passive microwave data are grouped into bins of size 0.1. The proposed model predictions at every pixel from each bin group was extracted and the mean and standard deviation were calculated. The mean for each bin was plotted along with its standard deviation in Figure 6.2. This evaluation was performed for the proposed model using NT2 and ASI estimates. From this figure, it can be seen that there is a strong bias towards high ice concentration values from the proposed model when compared with ASI estimates. This is appropriate because images selected for *Basic Train* were that are of good visual agreement with the SAR images, whereas images selected for *Hudson Strait Evaluation* were selected as images that required improvement (visual bias between ASI and SAR). Therefore this result, Figure 6.2b, is expected. A positive correlation can also be seen between the ASI and proposed model estimates. This shows that the proposed model and ASI estimates have similar relative predictions for ice concentration values. NT2 estimates are of coarser resolution due to the use of lower frequency channels, but they do not have a bias towards low ice concentration values. The proposed model has better agreement with the NT2 estimates for this dataset. The ASI estimates show a similar bias when compared with NT2. Note that although there is good agreement between the model and NT2, the proposed model captures fine details missed by the NT2 estimates.

Qualitative observations show that often the proposed model has the capability to predict small features that are missed by NT2 and ASI estimates. An example of the model successfully predicting boundaries in the marginal ice zone that NT2 and ASI estimates are incapable of doing is shown in Figure 6.3. The coarse resolution of passive microwave data prevents NT2 and ASI estimates from identifying the details in the marginal ice zone. Ice cracks are another example small features that NT2 and ASI estimates struggle identifying. As shown in Figure 6.4, NT2 estimates are unable to identify the ice cracks and ASI estimates identify them but do not capture the same level of details as the proposed model. The success of the proposed model in these instances are due to the leveraging of SAR data, which have higher spatial resolution than passive microwave data.



(a) (b) (c)

Figure 6.2: Comparison of estimates between (a) NT2 and ASI (b) proposed model and ASI (c) proposed model and NT2. Proposed model is biased towards higher ice concentrations when compared with ASI, which is expected given that the evaluation dataset corresponds to images for which the ice concentration was visibly underestimated. NT2 estimates show a similar bias when compared with ASI estimates. Proposed model does not show significant bias with NT2 estimates. The blue dots represent the mean of the estimates on the X and Y axes. The red line represents the standard deviation of the values for the Y axis. The dashed green line is $y = x$ line.



(a) (b) (c) (d)

Figure 6.3: (a) HH SAR image dated 01/20/2018 with marginal ice zone region shown in red rectangle. (b) ASI (c) NT2 and (d) proposed model ice concentration estimates overlaid on HH SAR image for region shown in rectangle. Proposed model captures more details that are missed by ASI and NT2 estimates due to the coarse resolution of the passive microwave sensors.

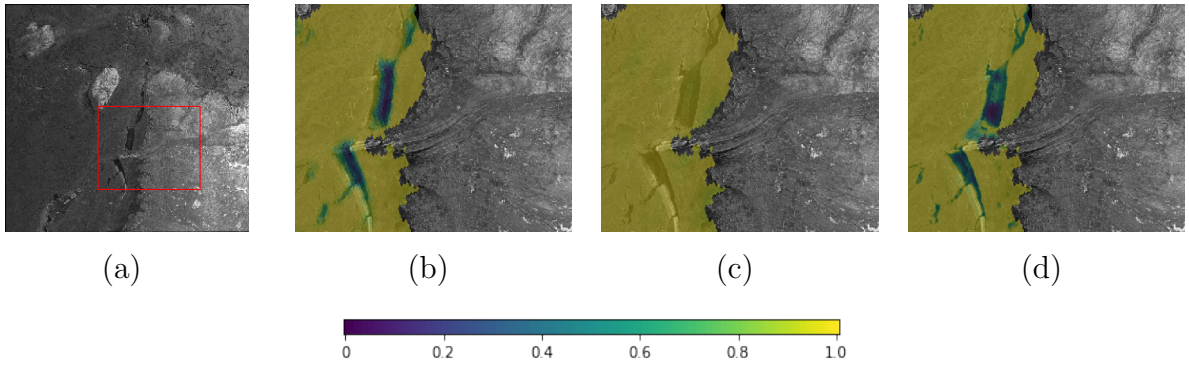


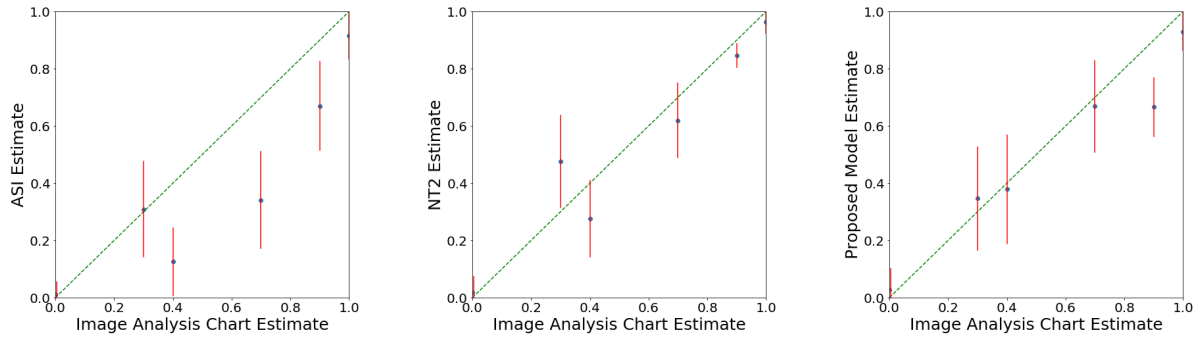
Figure 6.4: (a) HH SAR image dated 03/04/2018 with region consisting of ice cracks shown in red rectangle. (b) ASI (c) NT2 and (d) proposed model ice concentration estimates overlaid on HH SAR image for region shown in rectangle. Proposed model captures ice cracks more accurately than ASI estimates. NT2 estimates do not identify the ice cracks.

6.2 Image Analysis Chart Evaluation

Image analysis charts were also used to evaluate the model performance. The image analysis charts are not as numerically precise as passive microwave estimates or the proposed model predictions as ice concentration estimates from image analysis charts are only given in 0.1 intervals. To compare the results, the mean and standard deviation of the estimates from the pixels of each image analysis chart label value was obtained. The plots comparing the passive microwave estimates and proposed model predictions are shown in Figure 6.5. The lower ice concentration values of the ASI estimates in Figure 6.5a further demonstrates that the ASI estimates have bias. The proposed model and the NT2 estimates show better agreement with the image analysis charts. Statistics obtained from these plots are shown in Table 6.1. From this table, it can be seen that the proposed model and NT2 estimates have a lower mean absolute error and higher Pearson correlation coefficient than the ASI estimates when comparing with image analysis chart estimates. The proposed model also has slightly better agreement with image analysis charts than NT2.

The image analysis charts used in this chapter consisted of very few polygons for each ice concentration value. Therefore, it is of interest to look at how these results are representative of model performance. Standard error, which is defined as:

$$Standard\ Error = \frac{\sigma}{\sqrt{n}}, \tag{6.1}$$



(a) (b) (c)

Figure 6.5: Comparison of estimates between image analysis charts and (a) ASI, (b) NT2, and (c) proposed model estimates. ASI estimates have a bias towards lower ice concentration values when compared with image analysis charts. Proposed model and NT2 show better agreement with the image analysis chart estimates. The blue dots represent the mean of the estimates on the X and Y axes. The red line represents the standard deviation of the values for the Y axis. The dashed green line is $y = x$ line.

Table 6.1: Mean absolute error and Pearson correlation coefficient between ASI, NT2, and proposed model estimates with image analysis chart estimates obtained from plots from Figure 6.5

	ASI	NT2	Proposed Model
Mean Absolute Error	0.161	0.0812	0.0718
Pearson Correlation Coefficient	0.914	0.961	0.977

was used for this test. In this equation, σ represents the standard deviation of the sample set and n represents the number of samples.

For our test, the standard error of the mean of polygons was calculated. The steps to calculate this value are as follows:

1. The mean ice concentration value determined by ASI, NT2, and proposed model

estimates was calculated for each polygon from image analysis charts (polygon mean).

2. The standard deviation of the polygon means for each ice concentration value from image analysis charts was calculated. This is represented as σ in Equation 6.1.

3. Using n, which represents the number of polygons, and σ, which was calculated in step 2, the standard error is calculated.

Note the standard error calculated is different from the standard deviation that is shown in Figure 6.5, which is the standard deviation of estimates for ice concentration values of the image analysis charts. The standard error shows the spread of polygon means and is usually lower than the standard deviation because the spread of polygon means are smaller than the overall spread. The standard error calculation also provides equal weight to polygons, which means small polygons have the same weight as large polygons. This is different from the standard deviation calculation, where pixels are given the same weight and thus larger polygons have more impact. Calculating standard error for the polygon means gives insight on the likelihood of obtaining similar results when presented with a new polygon from same population.

The results of this test is shown in Table 6.2. For most observations, the standard error is low, which shows that the sample mean is close to the population mean.

Table 6.2: Standard error of the mean ice concentration value of ASI, NT2, and proposed model estimates for polygons from each ice concentration value on the image analysis charts.

Ice Concentration Value	Number of Polygons	ASI	NT2	Proposed Model
0.0	3	0.127	0.190	0.231
0.3	4	0.112	0.142	0.133
0.4	3	0.241	0.0368	0.0960
0.7	5	0.0753	0.0737	0.0981
0.9	3	0.0285	0.00148	0.00776
1.0	3	0.0321	0.0171	0.0335

Chapter 7

High Arctic Evaluation

In this chapter, the proposed model is evaluated on two images from the High Arctic. Not only was the model not exposed to the High Arctic, but the High Arctic is also very different from scenes that the model was exposed to. The High Arctic consists of more multi-year ice than the other datasets. This evaluation represents the scenario where the proposed model is used for a region that it was not exposed to and measures its ability to extend to new regions.

The images used for this study are dated 07/31/2013 and 09/11/2013 and the HH SAR images, ASI estimates, NT2 estimates, and proposed model predictions are shown in Figure 7.1. When comparing the HH SAR images from the High Arctic with those from the training data, some notable differences can be seen. There is more multi-year ice on these images, which is rare in the training sets (*Basic Train* and *Enhancement Set*). Open water of various tones can be seen in Figure 7.1e, which is a result of the varying incidence angles that alter the backscatter. The performance of the model in such situations have not been tested by the *Hudson Strait Evaluation* set. Moreover, the open water has a smoother appearance due to the lack of wind and ocean currents. Lastly, the overall appearance of the High Arctic SAR images are different that that from the training set. This is due to the very different surface conditions, such as thicker ice and smoother water, that have different backscatter patterns for the SAR signal.

For this evaluation procedure, the proposed model predictions are quantitatively compared with estimates retrieved from passive microwave data and Landsat data and qualitatively compared with estimates retrieved from passive microwave data.

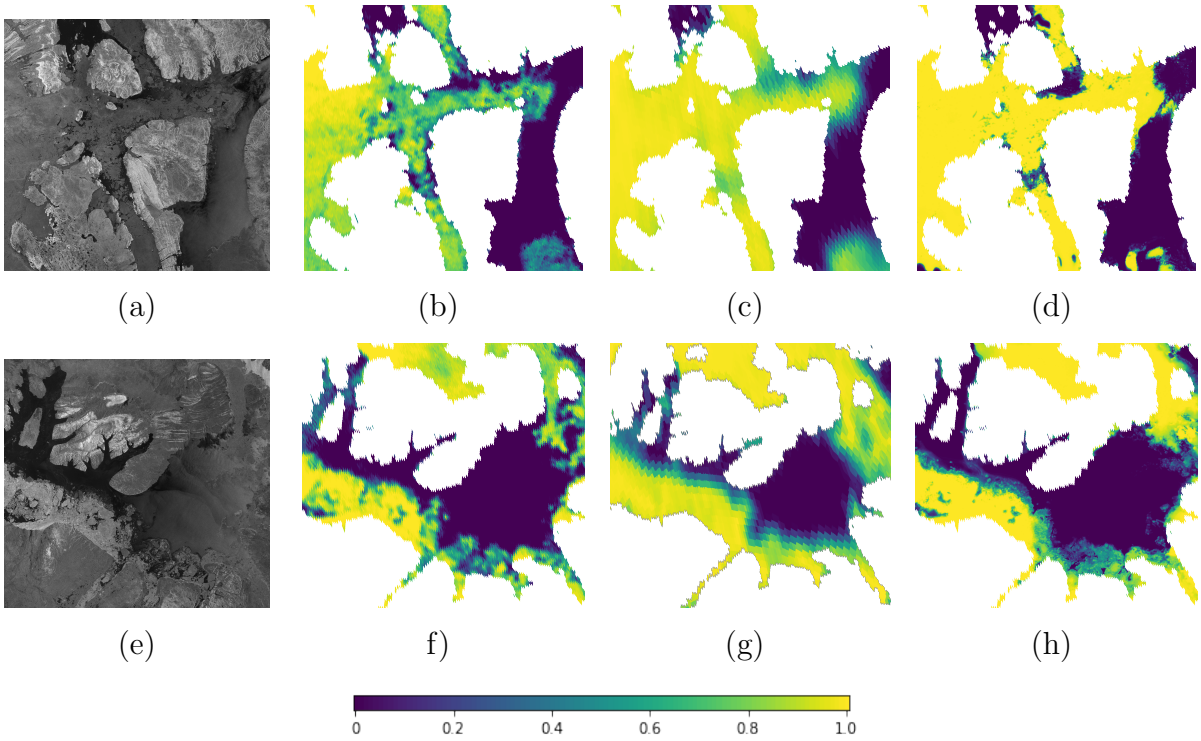


Figure 7.1: (a, e) HH SAR, (b, f) ASI estimates, (c, g) NT2 estimates, and (d, h) proposed model predictions for images dated 07/31/2013 (a, b, c, d) and 09/11/2013 (e, f, g, h).

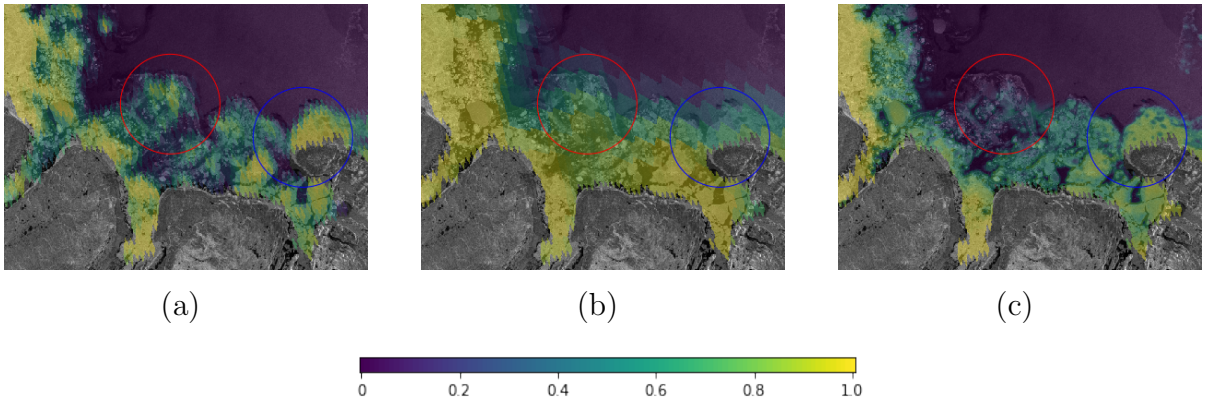
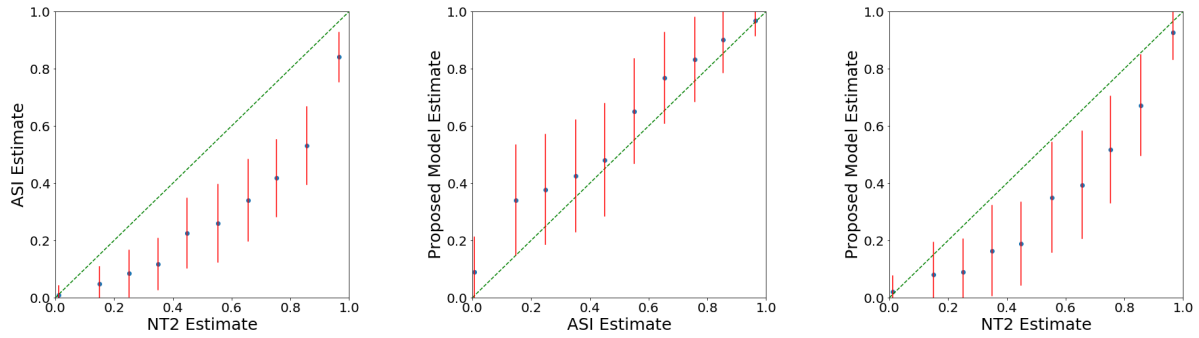


Figure 7.2: (a) ASI, (b) NT2, and (c) proposed model predictions overlaid on region of HH SAR image dated 09/11/2013. The proposed model misses features identified by ASI (red circle) while it captures fine details that ASI and NT2 estimates miss (blue circle). The coarse resolution of the NT2 estimates make it incapable of identifying fine details.

7.1 Qualitative Observations

From Figure 7.1, it can be seen that all estimates have room for improvement as details from both scenes are being missed. The proposed model captures details in the marginal ice zone that are missed by the ASI and NT2 estimates. An example of the proposed model capturing fine details missed by ASI and NT2 estimates and an example of the model missing features captured by ASI and NT2 estimates is shown in Figure 7.2. Furthermore, from Figure 7.1, it can be seen that the ASI estimates have a bias towards low ice concentration values as it predicts consolidated ice regions as marginal ice. It can also be seen on Figure 7.1 that both ASI and NT2 estimates have more noise within the open water region. The proposed model successfully identifies open water of different intensities on the SAR image accurately. On the top left area of Figure 7.1e, water appears dark. On the top right corner, water appears very bright. This is due to the effect of the incidence angle and the model labels both regions correctly as water. This suggests that the model learnt patterns of ice and water from the SAR image rather than relying on backscatter alone to determine ice concentration.



(a) (b) (c)

Figure 7.3: Comparison between (a) NT2 and ASI, (b) proposed model and ASI, (c) proposed model and NT2 estimates. Proposed model predictions are closer to ASI estimates than NT2 estimates. ASI estimates have a bias towards lower ice concentration values when compared to the model predictions and NT2 estimates. NT2 estimates have a bias towards higher ice concentration values when compared to ASI estimates and model predictions. The dashed green line is $y = x$ line.

7.2 Passive Microwave Evaluation

In this section, model predictions are compared with passive microwave estimates derived from ASI and NT2 data. The same procedure to compare estimates used in section 6.1 is used in this section. Results of the comparison are shown in Figure 7.3. From this figure, it can be seen that ASI estimates have a bias towards lower ice concentrations when compared to model predictions and NT2 estimates. This was also the case for the Hudson Strait evaluation in chapter 6 and visually apparent in Figure 7.1, as discussed in section 7.1. The NT2 estimates have a bias towards higher ice concentration values when compared with ASI estimates and model predictions. This is due to the NT2 estimates predicting marginal ice zones as ice and being unable to identify leads, due to the coarse resolution.

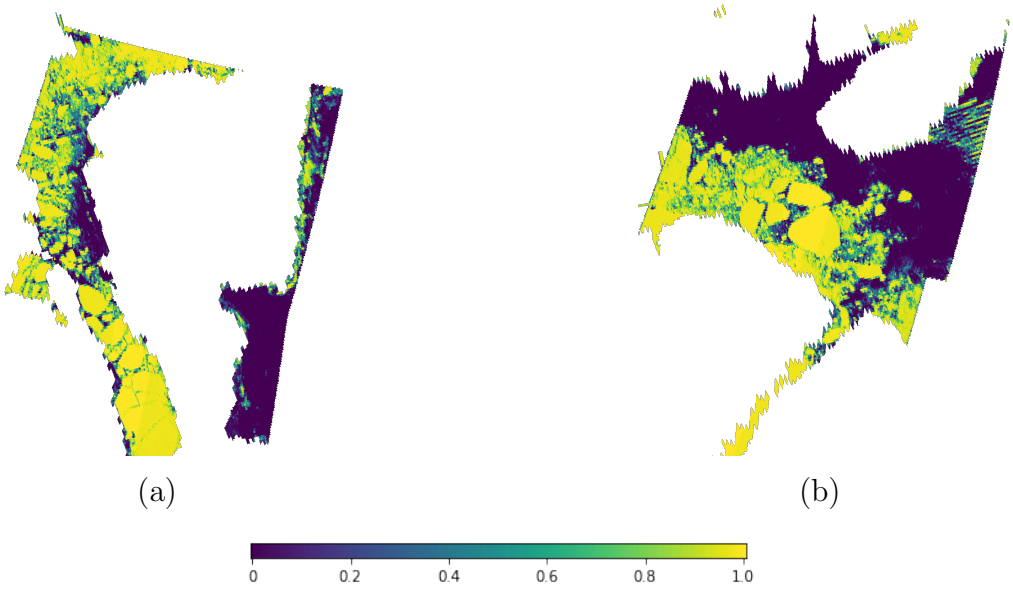
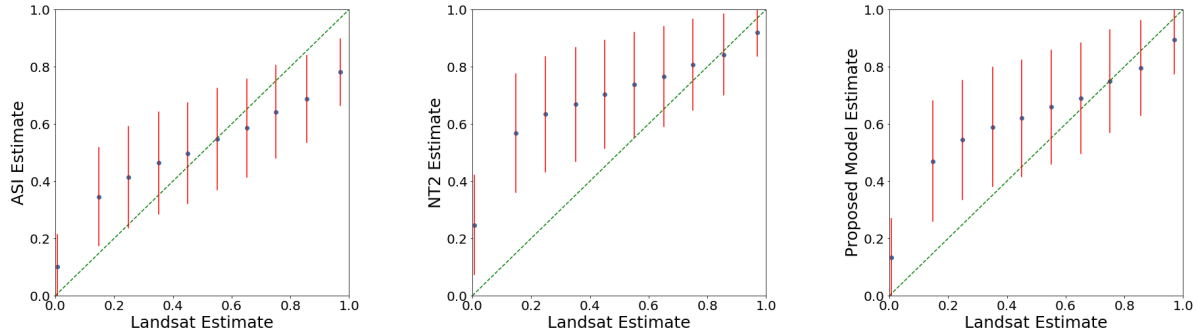


Figure 7.4: Regions for which Landsat estimates were acquired for images dated (a) 07/31/2013 and (b) 09/11/2013. Landsat estimates were only acquired for a small portion of original image.

7.3 Landsat Evaluation

In this section, we compare model predictions with Landsat estimates, which are of higher spatial resolution than passive microwave data based estimates and SAR based estimates. Landsat estimates were only acquired for a small region of the original SAR images; therefore the comparison does not represent the full image well. The region for which Landsat estimates were obtained is shown in Figure 7.4. From Figure 7.4, it can be seen that Landsat estimates, due to the higher spatial resolution, capture a lot of small details that estimates obtained from passive microwave are unable to.

The procedure for the Landsat evaluation is the same as that from section 6.1 and section 7.2. The results of this evaluation procedure is shown in Figure 7.5. It can be seen that the proposed model, ASI, and NT2 estimates do not have very good agreement with Landsat estimates. ASI estimates are predicting lower ice concentrations for regions Landsat estimates have as high ice concentration. Proposed model estimates are predicting higher ice concentrations for regions that Landsat estimates have as low ice concentration.



(a) (b) (c)

Figure 7.5: Comparison between Landsat estimates and (a) ASI (b) NT2 (c) proposed model estimates. ASI estimates and proposed model estimates show better agreement with Landsat estimates than NT2 estimates. The blue dots represent the mean of the estimates on the X and Y axes. The red line represents the standard deviation of the values for the Y axis. The dashed green line is $y = x$ line.

NT2 estimates are similar to the proposed model in predicting higher ice concentration values for regions that Landsat estimates have as low ice concentration, but NT2 estimates have an even stronger bias. Although, all figures in Figure 7.5 show a bias, they also show increasing means, which means that they have similar relative estimates as Landsat estimates.

The mean absolute error and Pearson correlation coefficient for ASI and NT2 estimates and model predictions are shown in Table 7.1. From this table, ASI estimates have the lowest mean absolute error and highest Pearson correlation coefficient, followed by the proposed model. This is due to the proposed model performing poorly in the region Landsat estimates were provided for the image dated 07/31/2013. This can be seen clearly when comparing the Landsat estimates from Figure 7.4a with predictions made by the proposed model in Figure 7.1c.

Table 7.1: Mean absolute error and Pearson correlation coefficient between ASI, NT2, and proposed model estimates with Landsat estimates obtained from plots from Figure 7.5

	ASI	NT2	Proposed Model
Mean Absolute Error	0.1151	0.2034	0.1434
Pearson Correlation Coefficient	0.9671	0.9164	0.9321

Chapter 8

Summary and Conclusion

8.1 Results Summary

In this thesis, a new method to obtain sea ice concentration estimates from HH and HV polarized SAR images is shown. A U-net, which accepts SAR images of any size to estimate ice concentration is used. This has significant benefits over the traditionally used CNNs for ice concentration estimation. Using a U-net, the runtime is significantly reduced when making predictions as the model does not need to be run on every pixel in the image. Furthermore, the U-net is not affected by the receptive field issue that traditional CNNs suffer from.

The U-net is trained with passive microwave data that has relatively coarse spatial resolution and may be biased in the marginal ice zone. Specifically, the model was trained with ice concentration estimates obtained from the ASI algorithm applied on passive microwave data of frequency 89 GHz (ASI estimates). It is shown that the issues of training with such estimates can be overcome by applying machine learning techniques. Through a 10-fold cross validation with the train set, a U-net using an L1 cost function, data augmentation, and curriculum learning was selected as the proposed model. From this test, the proposed model has a mean absolute error of 0.0387 with a standard deviation of 0.00387. The proposed model benefits from L1 loss function because L1 loss function is more robust to outliers than L2 loss function. Applying data augmentation improved model performance by reducing the error and eliminating bias that existed in the ASI estimates. Curriculum learning increased model stability and preserved more details in the marginal ice zones.

The proposed model was evaluated on unseen images from the Hudson Strait and High Arctic and compared with passive microwave data, image analysis charts, and Landsat

based estimates. The Hudson Strait images represent a scenario in which a model makes predictions on regions that it was exposed to while the High Arctic images represent a scenario where the model is extended to new regions with new ice types. For the passive microwave data comparison, ASI estimates were used along with another source that applies the NT2 algorithm on lower frequency channels (NT2 estimates). NT2 estimates have coarser resolution than the ASI estimates but they are not affected by atmospheric moisture to the same degree.

For the test images from the Hudson Strait, they were selected where the ASI estimates are visibly incorrect. The proposed model shows that it does not duplicate the ASI estimates and has better agreement with the NT2 estimates. Furthermore, the proposed model captures details in the marginal ice zone that are missed by ASI and NT2 estimates. From the comparison with image analysis charts, the proposed model performs better than ASI and NT2 estimates as it achieves a mean absolute error of 0.0718 compared with 0.0812 and 0.161 of NT2 and ASI estimates.

For the test images from the High Arctic, the proposed model performance was less impressive than its performance with the test images from the Hudson Strait. The proposed model captures fine details missed by ASI and NT2 estimates, but there is more regions that it predicts incorrectly. This suggests that the novel ice conditions, such as multi-year ice, are very different from the ice conditions the model has been exposed to. Furthermore, the overall SAR image appearance is different than that which the model was trained with.

In the final experiment, we attempt staged training to see if a deep learning model can benefit from using different training labels. In this experiment, we show that the deep learning model predictions contain select characteristics from the different training labels used to train it. We also found that the models trained with staged training has better performance than the models trained with each individual training label.

Ultimately, in this study, it is shown that deep learning with a U-net can be successful when retrieving ice concentration estimates, even when using inconsistent training labels, such as ASI estimates. Not only do we show a new method to obtain ice concentration estimates with passive microwave data, we also retain good estimates in consolidated ice and improve estimates in marginal ice zones.

8.2 Future Work

Due to the expansive realm of deep learning and the limitations of this study, there are opportunities for extensions to this research. Several aspects of this study can be explored

further to improve model performance.

The main limitation of this study is the limited data set. For example, the training set had a limited number of images with rough water conditions and multi-year ice and its effects were shown in the test set. A more thorough study can be performed with a larger, more inclusive set to develop a deep learning model that generalizes to many scene types. There was also a limited number of samples with marginal ice zone regions. This was the most problematic region and increasing the exposure to marginal ice zones can improve model performance. Staged training was also found to be advantageous in our experiment. A more thorough experiment could not be conducted because NT2 estimates were not available for the *Enhancement Set*. Predictions made from models using staged training may be more impressive if a larger training set was used. It would also be interesting to implement staged training with other types of training labels, such as image analysis chart or Landsat estimates. More thorough evaluations can also be conducted with a larger evaluation set. Due to the limited number of marginal ice zone regions, only 4 evaluation images from the Hudson Strait and 2 from the High Arctic were allocated for evaluation. A better understanding of model performance can be obtained with more evaluation images from different scenes. In chapter 7, less impressive performance was seen when the model was evaluated on images from the High Arctic. It would also be interesting to see model performance in other ice types, such as freshwater ice. Although properties of ice are different, contrast between consolidated ice and open water would be similar. Lastly, in this study, dual polarized SAR images are used. Each SAR channel provides different information. Therefore, more information can be retrieved from quad polarized SAR images, which can lead to improvements in model performance.

There are additional methods that can be applied to improve model performance, but it was outside the scope of this project. For example, SAR images contain banding and speckle noise that is problematic for deep learning models because they detect them as features. Although it is very difficult to remove this noise, there have been methods proposed to remove the noise and it can be beneficial to explore this realm [22, 68]. In recent years, there has been a lot of research in deep learning. State of the art deep learning models consist of many more layers than the ones implemented in this thesis. Larger models were not found to be advantageous, but this could be due to overfitting of our limited training set. This would be an avenue to explore if more data was acquired. Furthermore, there are also deep learning techniques, such as using an ensemble of deep learning models, that can improve performance.

References

[1] Rolf Adams and Leanne Bischof. Seeded region growing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(6):641–647, 1994.

[2] Søren Andersen, Rasmus Tonboe, Stefan Kern, and Harald Schyberg. Improved retrieval of sea ice total concentration from spaceborne passive microwave observations using numerical weather prediction model fields: An intercomparison of nine algorithms. *Remote Sensing of Environment*, 104(4):374–392, 2006.

[3] Jonathan Andrews, David Babb, and David Barber. Climate change and sea ice: Shipping in Hudson Bay, Hudson Strait, and Foxe Basin (1980–2016). *Elementa: Science of the Anthropocene*, 6(1), 2018.

[4] Matt Arkett, Dean Flett, Roger De Abreu, and Cameron Gillespie. Sea ice type and open water discrimination for operational ice monitoring with radarsat-2. In *2006 IEEE International Symposium on Geoscience and Remote Sensing*, pages 1631–1634. IEEE, 2006.

[5] Osiris A Valdez Banda, Floris Goerlandt, Jakub Montewka, and Pentti Kujala. A risk analysis of winter navigation in Finnish Sea areas. *Accident Analysis & Prevention*, 79:100–116, 2015.

[6] Yoshua Bengio, Jérôme Louradour, Ronan Collobert, and Jason Weston. Curriculum learning. In *Proceedings of the 26th Annual International Conference on Machine Learning*, pages 41–48, 2009.

[7] Bernhard E Boser, Isabelle M Guyon, and Vladimir N Vapnik. A training algorithm for optimal margin classifiers. In *Proceedings of the Fifth Annual Workshop on Computational Learning Theory*, pages 144–152, 1992.

[8] Lawson Brigham. Marine protection in the Arctic cannot wait. *Nature*, 478(7368):157–157, 2011.

[9] Canada. Transport Canada, GW Timco, B Gorman, J Falkingham, and B O'Connell. *Scoping Study: Ice Information Requirements for Marine Transportation of Natural Gas from the High Arctic*. National Research Council Canada., 2005.

[10] Frank D Carsey. *Microwave Remote Sensing of Sea Ice*. American Geophysical Union, 1992.

[11] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L Yuille. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(4):834–848, 2017.

[12] Liang-Chieh Chen, George Papandreou, Florian Schroff, and Hartwig Adam. Rethinking atrous convolution for semantic image segmentation. *arXiv preprint arXiv:1706.05587*, 2017.

[13] David A Clausi. Comparison and fusion of co-occurrence, Gabor and MRF texture features for classification of SAR sea-ice imagery. *Atmosphere-Ocean*, 39(3):183–194, 2001.

[14] Colin LV Cooke and K Andrea Scott. Estimating sea ice concentration from SAR: Training convolutional neural networks with passive microwave data. *IEEE Transactions on Geoscience and Remote Sensing*, 2019.

[15] Huawu Deng and David A Clausi. Unsupervised segmentation of synthetic aperture radar sea ice imagery using a novel Markov random field model. *IEEE Transactions on Geoscience and Remote Sensing*, 43(3):528–538, 2005.

[16] Alexandre S Gagnon and William A Gough. Trends in the dates of ice freeze-up and breakup over Hudson Bay, Canada. *Arctic*, pages 370–382, 2005.

[17] Xavier Glorot, Antoine Bordes, and Yoshua Bengio. Deep sparse rectifier neural networks. In Geoffrey Gordon, David Dunson, and Miroslav Dudík, editors, *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, volume 15 of *Proceedings of Machine Learning Research*, pages 315–323, Fort Lauderdale, FL, USA, 11–13 Apr 2011. PMLR.

[18] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT press, 2016.

[19] Government of Canada. Manual of ice (MANICE). 2016. https://www.canada.ca/en/environment-climate-change/services/weather-manuals-documentation/manice-manual-of-ice.html, [Online; accessed August 2020].

[20] Christopher Houser and William A Gough. Variations in sea ice in the Hudson Strait: 1971-1999. *Polar Geography*, 27(1):1–14, 2003.

[21] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. Densely connected convolutional networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4700–4708, 2017.

[22] Mahboob Iqbal, Jie Chen, Wei Yang, Pengbo Wang, and Bing Sun. Kalman filter for removal of scalloping and inter-scan banding in scansar images. *Progress In Electromagnetics Research*, 132:443–461, 2012.

[23] Natalia Ivanova, LT Pedersen, RT Tonboe, Stefan Kern, G Heygster, T Lavergne, A Sørensen, Roberto Saldo, G Dybkjær, L Brucker, et al. Inter-comparison and evaluation of sea ice algorithms: Towards further identification of challenges and optimal approach using passive microwave observations. *The Cryosphere*, 9:1797–1817, 2015.

[24] Jet Propulsion Laboratory. Get to know SAR. https://nisar.jpl.nasa.gov/mission/get-to-know-sar/overview/, [Online; accessed August 2020].

[25] Juha Karvonen. Baltic sea ice concentration estimation based on C-band HH-polarized SAR Data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 5(6):1874–1884, 2012.

[26] Juha Karvonen. Baltic sea ice concentration estimation based on C-band Dual-Polarized SAR data. *IEEE Transactions on Geoscience and Remote Sensing*, 52(9):5558–5566, 2014.

[27] Dilpreet Kaur and Yadwinder Kaur. Various image segmentation techniques: A review. *International Journal of Computer Science and Mobile Computing*, 3(5):809–814, 2014.

[28] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[29] Clark R.N. Swayze G.A. Livo K.E. Hoefen T.M. Pearson N.C. Wise R.A. Benzel W.M. Lowers H.A. Driscoll R.L. Kokaly, R.F. and A.J. Klein. USGS Spectral Library Version 7: U.S. Geological Survey Data Series 1035, 61 p. 2017. https://doi.org/10.3133/ds1035.

[30] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.

[31] Ruirui Li, Wenjie Liu, Lei Yang, Shihao Sun, Wei Hu, Fan Zhang, and Wei Li. Deepunet: A deep fully convolutional network for pixel-level sea-land segmentation. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 11(11):3954–3962, 2018.

[32] Andy Liaw, Matthew Wiener, et al. Classification and regression by randomforest. *R news*, 2(3):18–22, 2002.

[33] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3431–3440, 2015.

[34] William Lotter, Greg Sorensen, and David Cox. A multi-scale CNN and curriculum learning strategy for mammogram classification. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, pages 169–177. Springer, 2017.

[35] Junshen Lu, Georg Heygster, and Gunnar Spreen. Atmospheric correction of sea ice concentration retrieval for 89 GHz AMSR-E observations. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 11(5):1442–1457, 2018.

[36] Aleksey Malinka, Eleonora Zege, Georg Heygster, and Larysa Istomina. Reflective properties of white sea ice and snow. *Cryosphere*, 10(6), 2016.

[37] David Malmgren-Hansen, Leif Toudal Pedersen, Allan Aasbjerg Nielsen, Matilde Brandt Kreiner, Roberto Saldo, Henning Skriver, John Lavelle, Jørgen Buus-Hinkler, and Klaus Harnvig Krane. A convolutional neural network architecture for Sentinel-1 and AMSR2 data fusion. *IEEE Transactions on Geoscience and Remote Sensing*, 2020.

[38] Songrit Maneewongvatana and David M Mount. Analysis of approximate nearest neighbor searching with clustered point sets. *Data Structures, Near Neighbor Searches, and Methodology*, 59:105–123, 2002.

[39] Thorsten Markus and Donald J Cavalieri. The AMSR-E NT2 sea ice concentration algorithm: Its basis and implementation. *Journal of The Remote Sensing Society of Japan*, 29(1):216–225, 2009.

[40] Dimitrios Marmanis, Jan D Wegner, Silvano Galliani, Konrad Schindler, Mihai Datcu, and Uwe Stilla. Semantic segmentation of aerial images with an ensemble of CNNs. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 2016*, 3:473–480, 2016.

[41] T. Markus Meier, W. N. and J. C. Comiso. AMSR-E/AMSR2 unified l3 daily 12.5 km brightness temperatures, sea ice concentration, motion snow depth polar grids, version 1. [Hudson Strait, Labrador Sea, Gulf of St. Lawrence, Beaufort Sea, High Arctic]. Boulder, Colorado USA. NASA National Snow and Ice Data Center Distributed Active Archive Center. http://dx.doi.org/10.1029/2005JC003384 [Aug 2019].

[42] Mari-Ann Moen, Anthony Paul Doulgeris, Stian Normann Anfinsen, Angelika HH Renner, Nick Hughes, Sebastian Gerland, and Torbjørn Eltoft. Comparison of feature based segmentation of full polarimetric SAR satellite sea ice images with manually drawn ice charts. 2013.

[43] Alberto Moreira, Pau Prats-Iraola, Marwan Younis, Gerhard Krieger, Irena Hajnsek, and Konstantinos P Papathanassiou. A tutorial on synthetic aperture radar. *IEEE Geoscience and Remote Sensing Magazine*, 1(1):6–43, 2013.

[44] Olivia Mussells, Jackie Dawson, and Stephen Howell. Using RADARSAT to identify sea ice ridges and their implications for shipping in Canada's Hudson Strait. *Arctic*, pages 421–433, 2016.

[45] Vinod Nair and Geoffrey E Hinton. Rectified linear units improve restricted boltzmann machines. In *ICML*, 2010.

[46] NASA Earth Science Data Systems. What is synthetic aperture radar? Last modified 16 Apr 2020. https://earthdata.nasa.gov/learn/what-is-sar, [Online; accessed July 2020].

[47] National Center for Atmospheric Research Staff (Eds). The climate data guide: Sea ice concentration data: Overview, comparison table and graphs. Last modified 11 Sep 2017. https://climatedataguide.ucar.edu/climate-data/sea-ice-concentration-data-overview-comparison-table-and-graphs, [Online; accessed July 2020].

[48] National Oceanic and Atmospheric Administration. What is an eddy? 2020. https://oceanservice.noaa.gov/facts/eddy.html, [Online; accessed August 2020].

[49] National Resources Canada. Microwave remote sensing. Last modified 1 Dec 2015. https://www.nrcan.gc.ca/maps-tools-publications/satellite-imagery-air-photos/remote-sensing-tutorials/microwave-remote-sensing/9371, [Online; accessed August 2020].

[50] National Snow Ice Data Center. Thermodynamics: Albedo. Last modified 3 Apr 2020. https://nsidc.org/cryosphere/seaice/processes/albedo.html, [Online; accessed July 2020].

[51] National Snow and Ice Data Center. Cryosphere glossary: Sea ice. 2020. https://nsidc.org/cryosphere/glossary-terms/sea-ice, [Online; accessed August 2020].

[52] National Snow and Ice Data Center. Remote sensing: Passive microwave. Last modified 3 Apr 2020. https://nsidc.org/cryosphere/seaice/study/passive_remote_sensing.html, [Online; accessed July 2020].

[53] Kazuo Ouchi. Recent trend and advance of synthetic aperture radar with selected topics. *Remote Sensing*, 5(2):716–807, 2013.

[54] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Kopf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. Pytorch: An imperative style, high-performance deep learning library. In H. Wallach, H. Larochelle, A. Beygelzimer, F. dAlché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems 32*, pages 8024–8035. Curran Associates, Inc., 2019.

[55] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[56] Donald K Perovich et al. The optical properties of sea ice. 1996.

[57] Larissa Pizzolato, Stephen EL Howell, Chris Derksen, Jackie Dawson, and Luke Copland. Changing sea ice conditions and marine transportation activity in Canadian Arctic Waters between 1990 and 2012. *Climatic Change*, 123(2):161–173, 2014.

[58] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 234–241. Springer, 2015.

[59] Stein Sandven, Ola M Johannessen, and Kjell Kloster. Sea ice monitoring by remote sensing. *Encyclopedia of Analytical Chemistry: Applications, Theory and Instrumentation*, 2006.

[60] Nikolaos Sarafianos, Theodore Giannakopoulos, Christophoros Nikou, and Ioannis A Kakadiaris. Curriculum learning for multi-task classification of visual attributes. In *Proceedings of the IEEE International Conference on Computer Vision Workshops*, pages 2608–2615, 2017.

[61] Gunnar Spreen, Lars Kaleschke, and Georg Heygster. Sea ice remote sensing using AMSR-E 89-GHz channels. *Journal of Geophysical Research: Oceans*, 113(C2), 2008. http://dx.doi.org/10.1029/2005JC003384.

[62] Yuxing Tang, Xiaosong Wang, Adam P Harrison, Le Lu, Jing Xiao, and Ronald M Summers. Attention-guided curriculum learning for weakly supervised classification and localization of thoracic diseases on chest radiographs. In *International Workshop on Machine Learning in Medical Imaging*, pages 249–258. Springer, 2018.

[63] Lei Wang. Learning to estimate sea ice concentration from sar imagery. 2016.

[64] Lei Wang, K Scott, and David Clausi. Sea ice concentration estimation during freeze-up from SAR imagery using a convolutional neural network. *Remote Sensing*, 9(5):408, 2017.

[65] Yiru Wang, Weihao Gan, Jie Yang, Wei Wu, and Junjie Yan. Dynamic curriculum learning for imbalanced data classification. In *Proceedings of the IEEE international conference on computer vision*, pages 5017–5026, 2019.

[66] Daphna Weinshall, Gad Cohen, and Dan Amir. Curriculum learning by transfer learning: Theory and experiments with deep networks. *arXiv preprint arXiv:1802.03796*, 2018.

[67] Heidrun Wiebe, Georg Heygster, and Thorsten Markus. Comparison of the ASI ice concentration algorithm with Landsat-7 ETM+ and SAR Imagery. *IEEE transactions on Geoscience and Remote Sensing*, 47(9):3008–3015, 2009.

[68] Hua Xie, Leland E Pierce, and Fawwaz T Ulaby. Sar speckle reduction using wavelet denoising and markov random field modeling. *IEEE Transactions on Geoscience and Remote Sensing*, 40(10):2196–2212, 2002.

[69] Wei Yao, Zhigang Zeng, Cheng Lian, and Huiming Tang. Pixel-wise regression using u-net and its application on pansharpening. *Neurocomputing*, 312:364–371, 2018.

[70] Fisher Yu and Vladlen Koltun. Multi-scale context aggregation by dilated convolutions. *arXiv preprint arXiv:1511.07122*, 2015.

[71] Qiyao Yu and David A Clausi. SAR sea-ice image analysis based on iterative region growing using semantics. *IEEE Transactions on Geoscience and Remote Sensing*, 45(12):3919–3931, 2007.

[72] Qiyao Yu and David A Clausi. IRGS: Image segmentation using edge penalties and region growing. *IEEE transactions on Pattern Analysis and Machine Intelligence*, 30(12):2126–2139, 2008.

[73] Matthew D Zeiler, Dilip Krishnan, Graham W Taylor, and Rob Fergus. Deconvolutional networks. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 2528–2535. IEEE, 2010.

[74] Zhengxin Zhang, Qingjie Liu, and Yunhong Wang. Road extraction by deep residual u-net. *IEEE Geoscience and Remote Sensing Letters*, 15(5):749–753, 2018.